Pramana – journal of physics

Editor

R Nityananda
Raman Research Institute, Bangalore

Associate Editor

H R Krishnamurthy
Indian Institute of Science, Bangalore

Editorial Board

G S Agarwal, *Physical Research Laboratory, Ahmedabad*
V Balakrishnan, *Indian Institute of Technology, Madras*
J K Bhattacharjee, *Indian Assoc. for the Cultivation of Science, Calcutta*
D N Bose, *Indian Institute of Technology, Kharagpur*
B M Deb, *Panjab University, Chandigarh*
Rohini M Godbole, *Indian Institute of Science, Bangalore*
S Kailas, *Bhabha Atomic Research Centre, Bombay*
R K Kaul, *The Institute of Mathematical Sciences, Madras*
A V Khare, *Institute of Physics, Bhubaneswar*
T Padmanabhan, *Inter-Univ. Centre for Astronomy and Astrophysics, Pune*
R Ramaswamy, *Jawaharlal Nehru University, New Delhi*
A K Raychaudhuri, *Indian Institute of Science, Bangalore*
K C Rustagi, *Centre for Advanced Technology, Indore*
E V Sampathkumaran, *Tata Institute of Fundamental Research, Bombay*
Abhijit Sen, *Institute for Plasma Research, Gandhinagar*
S K Sikka, *Bhabha Atomic Research Centre, Bombay*
Y Singh, *Banaras Hindu University, Varanasi*

Editor of Publications of the Academy

V K Gaur
C-MMACS, NAL, Bangalore

Annual Subscription Rates

All countries except India (Price includes AIR MAIL charges)	US$ 200
India	Rs. 200

Annual subscriptions for **individuals** for India and abroad are Rs. 100/- and $50 respectively.

All correspondence regarding subscription should be addressed to the **Circulation Department** of the Academy

Editorial Office

Indian Academy of Sciences, C V Raman Avenue,
P.B. No. 8005, Bangalore 560 080, India

Telephone: 334 2546
Telex: 845-2178 ACAD IN
Telefax: 91-80-334 6094

Contents – Parts I & II

Editors' note 1

Preface 3

Mathematical aspects of dynamical systems

Random matrices and matrix models: The JNU lectures .. *Madan Lal Mehta* 7

Hölder exponents of irregular signals and local fractional derivatives .... *Kiran M Kolwankar and Anil D Gangal* 49

Signature of chaos in power spectrum .... *M C Valsakumar, S V M Satyanarayana and V Sridhar* 69

Dynamics of simple one-dimensional maps under perturbation .... *Somdatta Sinha and Parichay K Das* 87

Chaos in modulated logistic systems .... *V M Nandakumaran* 99

Effect of boundary conditions on the invariant density of noisy maps at fully-developed chaos .... *V Balakrishnan, G Nicolis and C Nicolis* 109

Symplectic integration of nonlinear Hamiltonian systems .... *Govindan Rangarajan* 129

Integrable systems and solitons

Painlevé analysis and integrability aspects of some nonlinear partial differential equations .... *K Porsezian* 143

Localized coherent structures of $(2+1)$ dimensional generalizations of soliton systems .... *M Lakshmanan and R Radha* 163

Geometry and nonlinear evolution equations .... *Radha Balakrishnan* 189

Nonlinear chiral models: Soliton solutions and spatio-temporal chaos .... *M S Sriram and J Segar* 205

Spatio-temporal chaos, synchronization and control

Control methods for problems of mixing and coherence in chaotic maps and flows .... *Ashutosh Sharma and Neelima Gupte* 231

Characterization and control of chaotic dynamics in a nerve conduction model equation .... *S Rajasekar* 249

Control of chaos .... *Sangeeta D Gadre and V S Varma* 259

Effect of noise on coupled chaotic systems .... *Manojit Roy and R E Amritkar* 271

Turbulence

Some recent advances in the theory of homogeneous isotropic turbulence… …*Sujan K Dhar, Anirban Sain, Ashwin Pande and Rahul Pandit* 325

The randomly stirred fluid – Turbulence as a problem in statistical mechanics …*J K Bhattacharjee* 365

Quantum aspects of chaos

Chaos in the hydrogen atom interacting with external fields… …*K Ganesan and R Gębarowski* 379

Resonances and chaos in the collinear collision system (He, H_2^+) and its isotopic variants…*Susanta Mahapatra, N Sathyamurthy and Ramakrishna Ramaswamy* 411

Chaos in intrinsic motion of nuclei and its role in dissipative collective dynamics…*Santanu Pal* 425

Chaos and localization in coupled quartic oscillators… …*M S Santhanam, V B Sheorey and A Lakshminarayan* 439

Experiments on quantum chaos using microwave cavities: Results for the pseudo-integrable L-billiard…*A Kudrolli and S Sridhar* 459

Quantum chaology: The photoeffect and beyond…*Bala Sundaram* 469

Periodic orbits in polygonal billiards…*Debabrata Biswas* 487

Semiclassical theory for many-body fermionic systems… …*Pierre Gaspard and Sudhir R Jain* 503

Accuracy of trace formulas…*Arul Lakshminarayan* 517

The phase of the Riemann zeta function…*Avinash Khare* 537

Exotic spectra, wavefunctions and transport in incommensurate lattices… …*A Lahiri, T K Roy and A Bhowal* 555

Fractal characteristics of critical and localized states in incommensurate quantum systems…*Indubala I Satija and Jukka A Ketoja* 589

Applications

Dynamical signatures of 'phase transitions': Chaos in finite clusters… …*Vishal Mehra, Saroj K Nayak and Ramakrishna Ramaswamy* 603

Classical gauge theories as dynamical systems — Regularity and chaos… …*S Lakshmibala, Bindu A Bambah, M S Sriram and C Mukku* 617

..............................*S Sengupta, J R Bhatt, P K Kaw and J C Parikh* 655

Nonlinear propagation of relativistically intense electromagnetic waves in a collisionless plasma..*P K Kaw and A Sen* 675

The onset of chaos in the wake of an oscillating cylinder: Experiment and the dynamics of the circle map..
..............................*D J Olinger, A B Chhabra and K R Sreenivasan* 693

Chaos in jerky flow—Experimental verification of a theoretical prediction
....*S J Noronha, G Ananthakrishna,*
L Quaouire and C Fressengeas 705

Chaos and predictability of the Indian summer monsoon......*B N Goswami* 719

Some recent applications of cell dynamical modelling to phase ordering dynamics..*Sanjay Puri* 737

Noisy nonlinear coupled equations—Some new insights......................
.... *J K Bhattacharjee, Anita Mehta and J M Luck* 749

Nonlinear fluctuating hydrodynamics and sequence of time scales of relaxation in supercooled liquids..*Shankar P Das* 759

Editors' note

This special issue of *Pramana* on "Nonlinearity and chaos in the physical sciences" differs in some ways from other recent special issues we have published.

First, it is not the proceedings of a conference or a meeting but has resulted from a "virtual conference" involving the members of the Indian community working in the area and their friends and collaborators. The response has been greater than we anticipated and the credit is due to the guest editors, Profs R Ramaswamy and K R Sreenivasan, the contributors, and of course the vigorous state of the subject itself.

A second difference is that *Pramana* is not bringing out a separate supplement this time, but presenting the material as the January and February issues. While this split has to be made for practical reasons, the two parts should be regarded as one unit and we are indeed reproducing this note, the guest editors' preface, and the full table of contents in both the parts. Given the amount of work involved in proof-reading and copy editing (even with many of the manuscripts in electronic form) it would have been impractical to bring out separate January and February issues as well. The extra load has in fact meant a delay in the appearance of this two-part special issue which we sincerely regret. We hope to be better prepared when such occasions come up in the future.

This special issue reflects the current active state of the subject of nonlinear phenomena in physics and allied disciplines in the country. We hope that it will be read with great interest by followers and practitioners of this rapidly expanding area and stimulate further work. We also hope that many of the authors who contributed to this special issue will submit some of their work to *Pramana–J. Phys.* in future. And finally, there is no reason why other areas in which there is a strong community cannot follow the example set here.

It is a pleasure to thank Profs K R Sreenivasan and R Ramaswamy for their efforts which have helped this journal bring out an extended special issue devoted to an area of great current interest to physicists all over the world.

R Nityananda
H R Krishnamurthy

Preface

One of the most important conceptual advances of the modern era has been the realization that the temporal evolution of many nonlinear systems, independent of their physical origin, is profoundly similar. The similarity is dynamical and quantitative, not merely statistical and descriptive. The subject that cultivates and explores—to some degree or another—this universality of nonlinear problems is called nonlinear science, or nonlinear dynamics, or chaos, depending to some extent on the orientation of the researcher. Initial progress in the area was rapid and simultaneously occurred in several disciplines in the natural and mathematical sciences. As we have begun to appreciate better the richness of nonlinear systems, the challenges have proliferated as well.

From the time of inception of this multidisciplinary subject, several groups of scientists in India have contributed to it. These groups were initially scattered with no serious contact among them. In 1987, a summer school on the subject was organized in Bangalore, under the auspices of the Department of Science and Technology, with the intention of providing an impetus for further research and for developing a sense of community within the country. Many bright and young scientists participated with excitement and eagerness. Some of the seeds sown at the time have now matured, and the nonlinear science community seems to have established an identity of its own.

It was felt, ten years later, that a collection—in a single place—of the current work in the field would spur further synergism and collaborations. Thus came the inspiration for this special volume. The scope was restricted, by and large, to the physical sciences. Various researchers active in the field within the country (and some living abroad with strong ties to local activities) were invited to contribute to the volume. Given the special nature of the volume, they were encouraged to include in their contributions both review elements and new results. The enthusiasm with which the proposal was greeted by them is a testimony to both the interest that the field commands and the way in which the idea resonated with their own thinking. The articles were all subjected to peer review. While the papers are representative of the authors' specific interests and individually provide a glimpse of their own area of inquiry, it is our belief that the collection provides a comprehensive account of the work occurring in the field as a whole.

Given the vast scope of the work that this volume represents, it is difficult to pigeonhole it into a few neat categories but an attempt—even if imperfect—seems worthwhile. The forty two contributions have been grouped in six sections as follows.

A mathematical tool that is much used in nonlinear science, especially in quantum chaos and increasingly in other sub-fields of dynamical systems as well, is the theory of random matrices. Mehta's article, which begins this volume as well as Section I on mathematical aspects, is a comprehensive survey and summary of the most important results in this area. Other articles in the section by Balakrishnan, Nicolis and Nicolis, by Nandakumaran, and by Sinha and Das deal with iterated unimodal maps which have been important to the mathematical understanding of chaotic dynamics. Kolwankar and Gangal discuss the connection between fractional differentiability and the fractal dimension of curves; this topic is of considerable interest because of the ubiquitous presence of fractals in chaotic systems. The meaning of the power spectrum and its generic features form the focus of the article by Valsakumar, Satyanarayana and Sridhar, who are concerned with signal processing and data analysis. Finally, Rangarajan deals with a new technique for integrating the equations of motion of a Hamiltonian system while explicitly respecting its symplectic property.

Dynamical systems that are exactly integrable—and therefore nonchaotic—have been studied in great detail well before the spurt of activity in chaos theory. Integrable systems are constrained by independent conserved quantities, and support coherent solutions such as solitons. The properties of solitons and their many applications form the focus of Section II which contains articles by Porsezian on the Painlevé property, by Lakshmanan and Radha on solitons in higher dimensions, and by Balakrishnan on the geometric aspects of nonlinear evolution equations. An area where the concepts of solitons has seen much application is that of particle physics and field theory, and the article by Sriram and Segar discusses some of these issues in the context of nonlinear chiral models.

An aspect of chaotic systems that has been at the forefront of recent research is their control. Several groups in the country are involved in devising, implementing and studying control methods. Articles in this area are grouped in Section 3. The papers by Gadre and Varma, Sharma and Gupte, and Rajasekhar are representative of the work in this area. Related concerns deal with extended dynamical systems, namely coupled lattice systems, where spatiotemporal chaos is of interest. Articles by Sinha, by Parekh, Ravi Kumar and Kulkarni, and by Roy and Amritkar are on the latter topic.

While temporal chaos has unearthed exceedingly interesting behaviours of nonlinear systems, their connections to fluid turbulence are still tenuous. In the background however, there exists the hope that the knowledge one acquires from simpler spatiotemporal systems and the methods developed to probe them will enhance our ability to understand the richness of fully developed turbulence. Section IV contains two articles that deal with aspects of turbulence from a relatively modern perspective: that by Bhattacharjee deals with statistical mechanical aspects of the randomly stirred fluid, while the paper by Dhar, Sain, Pande and Pandit deals with developments in the scaling phenomenology, and emphasizes the lessons learnt from the GOY model.

The field of quantum chaos attempts to understand the difference in behaviour between systems that are (in some appropriate limit) classically chaotic and those that are classically integrable. There are, in contrast to classical chaos, a multitude of definitions of what constitutes quantum chaotic behaviour. One very successful classification of systems is based on the different universality classes of random matrices. Papers in Section V have analysed problems in quantum chaos from a variety of points of view. Ganesan and Gebarowski look at chaos in the hydrogen atom in different external fields, while Santhanam, Sheorey and Lakshminarayan study the morphology and characteristics of eigenstates of quartic oscillators; the paper by Mahapatra, Sathyamurthy and Ramaswamy analyzes chaos and resonances in a molecular collision system, while Kudrolli and Sridhar report on results of an experiment to study quantum chaos in a microwave cavity. Pal examines chaos in nuclear motion and its role in the collective dynamics, using results from a quantum chaotic analysis in a classical model for dissipation in nuclear fission. Periodic orbits have played a major role in semiclassical theories, and the papers by Biswas, by Lakshminarayan, by Gaspard and Jain and by Khare deal with different aspects of trace formulae. The similarity between the

The final section, broadly classified as **Applications**, shows why chaos has enjoyed universal appeal: rarely, it seemed, had the methods developed in *modern* mathematics found such an immediate relevance in real-life, although one often had to look for the connections rather meticulously: Chaos is in the details!

Mehra, Nayak and Ramaswamy apply the methods of nonlinear dynamics to understand the mesoscopic behavior of clusters of a small number of atoms. Lakshmibala, Bambah, Sriram and Mukku review studies of dynamical chaos in classical gauge theories using both analytic as well as numerical techniques. Ambika discusses the complex behaviour of Josephson junctions, such as its phase locked and chaotic states, scaling near basin boundaries, and so forth. The classical chaos elements of collective thermalization of the quark-gluon problem are discussed by Sengupta, Bhatt, Kaw and Parikh. The paper by Kaw and Sen, which deals with the nonlinear propagation of electromagnetic waves in plasmas, is concerned with applications to a variety of current experiments in laser fusion and plasma based accelerators. Olinger, Chhabra and Sreenivasan describe experimental results in the wake of a transversely oscillating cylinder in a fluid flow, which, they demonstrate, belongs to the same universality class as circle maps with cubic nonlinearity. The low-dimensional chaotic behavior of the stick-slip phenomenon common in plastic instability of flowing materials is discussed by Noronha, Ananthakrishna, Quaouire and Fressengeas. Goswami dwells on the limits to predictability of Indian monsoons due to the phase and amplitude irregularities of interseasonal oscillations in a dynamic model. Puri discusses phase ordering of random magnets and phase separation in binary fluids, while Das treats the nonlinear dynamics of fluctuations and relaxation times in supercooled liquids. Finally, Bhattacharjee, Mehta and Luck discuss the role of infrared divergences in a perturbation theory of noisy coupled maps.

As this summary suggests, the richness and variety of the subject and its practice in the country, which this collection reflects, is immense. The subject brings together, in a unique way, a unification of mathematics, fundamental physics in several of its branches, as well as applications. We have greatly enjoyed putting the volume together, and hope that you, the reader, will share our sense of discovery and delight.

However, none of this could have happened without the support and cooperation of the nonlinear science community. As authors and referees, they have played the vital and crucial role in giving this volume its present shape and they deserve our warmest thanks.

Ramakrishna Ramaswamy, Jawaharlal Nehru University, New Delhi.
Katepalli R. Sreenivasan, Yale University, New Haven.

Random matrices and matrix models: The JNU lectures

MADAN LAL MEHTA
School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067, India
Permanent Address: Centre d'Études Nucleaires de Saclay, 91191 Gif-sur-Yvette Cedex, France

Abstract. A course of lectures was given at the Jawaharlal Nehru University and the Jamia Milia Islamia, New Delhi, during February–March 1996. The following notes were distributed to the audience before each lecture. These notes, which are sketchy and do not go in details, were meant to help students follow the standard literature on the subject. They are collected here (with the exercises!) in the hope that they might prove useful to a larger community of research workers.

Keywords. Random matrices; matrix ensembles; statistical properties of eigenvalue spectra.

PACS Nos 02.10; 5.45; 5.90

1. Introduction

Before one can consider the problem of "random matrices" one has to know a few preliminary things.

- Numbers: A set in which addition, subtraction, multiplication and division can be performed with the usual rules of arithmetic, namely
  (a) addition is commutative and associative: $x+y=y+x$, $x+(y+z)=(x+y)+z$;
  (b) multiplication is associative: $x(yz)=(xy)z$;
  (c) existence of a zero and of a unit: $x+0=0+x=x$; $1.x=x.1=x$;
  (d) multiplication is distributive over addition: $x(y+z)=xy+xz$, $(y+z)x=yx+zx$;
  (e) existence of a negative and an inverse: $x+(-x)=(-x)+x=0$; $x\cdot x^{-1}=x^{-1}\cdot x=1$ if $x\neq 0$;

  is called a number field, its elements are numbers.
- There are only three kinds of numbers: real, complex and quaternion (theorem of Frobenius).
- Real and complex numbers—are very familiar: the complex numbers are formed with two real numbers and one special unit i with $i^2=-1$.
- Quaternions are not so familiar: they are formed with four real numbers and three units e_1, e_2 and e_3 with $e_1^2=e_2^2=e_3^2=e_1e_2e_3=-1$.
- Multiplication is commutative for real and complex numbers, but not for quaternions $xy\neq yx$.
- Matrices are rectangular arrays of numbers. With a square matrix one can usually associate a number called its determinant. Matrices can be added or multiplied if their

number of rows and columns are appropriate. The values of λ for which $\det(A - \lambda I)$ is zero, are called eigenvalues of A. (Here I is the unit matrix). For an eigenvalue λ of A the non-zero column matrix x satisfying the equation $Ax = \lambda x$ is called the eigenvector of A corresponding to the eigenvalue λ.

- A square matrix A of order $n \times n$ has n eigenvalues and at most n eigenvectors.
- One can define many elementary operations on matrices such as transposition, hermitian conjugation and taking the dual, the result is denoted respectively by A^T, $A^\dagger$ and A^D. We call A symmetric if $A = A^T$, hermitian if $A = A^\dagger$, self-dual if $A = A^D$, orthogonal if $AA^T = I$, unitary if $AA^\dagger = I$ and symplectic if $AA^D = I$.
- The eigenvalues of a real symmetric matrix, of a complex hermitian matrix and of a quaternion self-dual matrix are all real; those of a real orthogonal matrix, of a complex unitary matrix and of a quaternion symplectic matrix are all of the form $e^{i\theta}$, θ real.
- The eigenvalues of an arbitrary real or complex matrix are in general complex numbers, and the question of the eigenvalues of an arbitrary quaternion matrix is meaningless, since the determinant of such a matrix can not be defined in a reasonable way.
- The problem of *random matrices* is very simple to state. Given an $n \times n$ real symmetric random matrix (random means that the probability density of the matrix elements is given), what can one say about its eigenvalues? or about a few of its eigenvalues? or of its eigenvectors? or a few of its eigenvectors? The words "real symmetric" in the above sentence can be replaced by any of the following, "complex hermitian", "quaternion self-dual", "symmetric unitary", "unitary", "self-dual unitary", "real" or "complex".
- Why should one worry about such questions? Some history: In the 1920's and 1930's some mathematicians studied such questions in relation to their applications in preparing birth and death statistics for insurance companies. In 1942–1944, Selberg evaluated a related integral and used it to find the density and the distribution of prime numbers. But this paper remained unnoticed for about 40 years. More about it later. In the 1940's and 1950's nuclear bombs and nuclear power stations appeared, and for obvious reasons people wanted to know about the neutron resonance energies of many nuclei such as uranium, cadmium and aluminium; their positions, heights and widths. According to quantum mechanics each nucleus is described by a hamiltonian function, its eigenvalues giving the positions and widths of the resonances. So choose a base, express the hamiltonian as a large matrix in this base and compute its eigenvalues. The hamiltonian function, i.e. the nuclear interactions are not known so well, and even if known are quite complicated. Hence Wigner, who knew the question very well, suggested that for statistical properties it will be sufficient to take the hamiltonian matrix elements as random numbers. In the 1960's and 1970's a large amount of experimental data about nuclear resonances in various nuclei was collected; also the mathematical problem of random matrices was more or less solved for one particular model, namely when the matrix elements have a normal or Gaussian distribution. A comparison of the two was quite satisfactory. In the 1970's it was discovered somewhat unexpectedly that the zeros of the Riemann zeta function on the critical line behave as if they were the eigenvalues of a complex hermitian random matrix. (A small

concrete beams, the electromagnetic resonances in closed cavities of random shape, the possible quantum energies of a particle confined in a box of random shape, the positions of trees in a forest, etc. all seem to follow the same statistical laws as the eigenvalues of a random matrix. The mathematical methods developed for the random matrix problem find use in other fields such as quantum gravity and string theory.

A list of references is given at the end of these notes in a general bibliography. [1] and [2] are cited in the text as RM and MT respectively.

2. (Gaussian) Ensembles of matrices

2.1 *Invariance under a change of basis*

In these lectures I will consider ensembles of matrices, H, whose elements are random (explained below), and which are invariant under a transformation U of the basis, namely

$$P(UHU^{-1}) = P(H). \tag{1}$$

The three cases of importance are when H is

- real symmetric. U is then real orthogonal. This case has the parameter (to be explained below) $\beta = 1$.
- complex hermitian. U is then complex unitary and $\beta = 2$.
- quaternion self-dual. This case is unfamiliar, and the following subsection deals with this case in detail. U is then quaternion real symplectic and $\beta = 4$.

Here β is the number of real components of the matrix elements, 1 for real, 2 for complex and 4 for quaternion. Invariance implies that $P(H)$ depends only on the traces of powers of H.

$$P(H) = P(\mathrm{tr}\, H, \mathrm{tr}\, H^2, ..., \mathrm{tr}\, H^n). \tag{2}$$

(Exercise: Why only the first n powers? Hint: Only symmetric functions of the eigenvalues survive such transformations.)

2.2 *Digression on quaternions*

Quaternions (or quaternion numbers) are not so familiar as the real or complex numbers. Just as complex numbers have two components, quaternions have four components, $a = a_0 + a_1e_1 + a_2e_2 + a_3e_3 = a_0 + \vec{a}.\vec{e}$. One calls a_0 the scalar part, and $\vec{a}$ the vectorial part of a. The dual of a is $a^D = a_0 - \vec{a}.\vec{e}$. The units e_1, e_2, e_3 have the multiplication laws

$$e_1^2 = e_2^2 = e_3^2 = -1, \quad e_1e_2 = -e_2e_1 = e_3, \quad e_2e_3 = -e_3e_2 = e_1,$$
$$e_3e_1 = -e_1e_3 = e_2. \tag{3}$$

The three 2×2 Pauli matrices σ_1, σ_2 and σ_3 familiar to physicists, also have simila multiplication laws and therefore it is sometimes convenient to represent the e_j and th quaternions as 2×2 matrices,

$$C(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C(e_1) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad C(e_2) = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad C(e_3) = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

(4

$$C(a) = \begin{bmatrix} a_0 + ia_3 & a_1 + ia_2 \\ -a_1 + ia_2 & a_0 - ia_3 \end{bmatrix}, \quad C(a^D) = \begin{bmatrix} a_0 - ia_3 & -a_1 - ia_2 \\ a_1 - ia_2 & a_0 + ia_3 \end{bmatrix}.$$

(5

Thus an $n \times n$ quaternion matrix will sometimes be represented by a $2n \times 2n$ comple matrix. Usually the a_j are taken as real numbers, but we will allow them to be comple Thus some non-zero quaternions will not have an inverse. This being admitted, an $2n \times 2n$ complex matrix can be thought of as cut into 2×2 blocks and each 2×2 bloc expressed as a quaternion; if these quaternions are real (i.e. their components are re numbers), we call it a quaternion real matrix. This establishes a one to on correspondence between an $n \times n$ quaternion matrix A and a $2n \times 2n$ complex matri $C(A)$. One verifies easily (please do it!) that

$$C(A^D) = -C(e_1 I) C(A)^T C(e_1 I).$$

(6

Hence A is self-dual if and only if $C(e_1 I) C(A)$ is an anti-symmetric matrix of twice th size.

We state here one definition and two theorems without proof.

Theorem 1. *For a quaternion matrix A it is impossible to define a determinant havi the three properties:*

(i) $\det A = 0$, *if and only if $Ax = 0$ has a non-zero solution x;*
(ii) $\det(AB) = \det A \cdot \det B$;
(iii) $\det A$ *is multilinear in the rows (or in the columns) of A.*

(For a proof see Chapter 8 of MT.) In order to define a determinant of a quaternion matr one has to give up one or more of the above properties.

DEFINITION 1

For a quaternion matrix A we define

$$\det A = \sum_P (-1)^P (a_{j_1 j_2} a_{j_2 j_3} \cdots a_{j_r j_1})_0 (a_{k_1 k_2} a_{k_2 k_3} \cdots a_{k_s k_1})_0 \cdots$$

(

where the sum is over all permutations P of the indices $(1, 2, \ldots, n)$ consisting of t exclusive cycles $(j_1, j_2, \ldots, j_r)$, $(k_1, k_2, \ldots, k_s), \ldots$. The subscript 0 means that o should take the scalar part of each cyclic product. If the quaternion real matrix A is se dual, then the subscript 0 in eq. (7) above is not necessary under certain conditions. (s Chapter 8 of MT for details). (Exercise: Which of the three properties of determinants missing in this definition?).

Theorem 2. *Let $C(A)$ be the $2n \times 2n$ complex matrix corresponding to the $n \times n$ self-dual quaternion matrix A. Then $C(e_1 I)C(A)$ is a $2n \times 2n$ anti-symmetric matrix and the determinant of A is equal to the Pfaffian of $C(e_1 I)C(A)$*

$$\det A = \mathrm{Pf}[C(e_1 I)C(A)]. \tag{8}$$

(For a proof see Chapter 8 of MT, or devise one by yourself.)

2.3 *Statistical independence of the various elements*

$$P(H) = \prod_j f_j(H_{jj}) \prod_{j<k} \prod_{\alpha=0}^{\beta-1} f_{jk\alpha}(H_{jk}^{(\alpha)}). \tag{9}$$

These requirements then force $P(H)$ to be an exponential with only first two powers of H.

$$P(H) = \exp(-a \operatorname{tr} H^2 + b \operatorname{tr} H + c). \tag{10}$$

For a proof consider $H = U^{-1}H'U$, with $U - I$ having non-zero small elements ε and $-\varepsilon$ only in the position (1,2) and (2,1) respectively. Then compare $P(H)$ and $P(H')$ to the first order in ε. For the anti-symmetric real, anti-hermitian complex or anti-self-dual quaternion real matrices, the same two constraints, namely (i) invariance under an orthogonal real, unitary complex or symplectic quaternion real matrix, and (ii) statistical independence of linearly independent various real parameters entering in the matrix elements, lead to $P(H) = \exp[-a \operatorname{tr} H^2 + c]$.

2.4 *Change of variables to eigenvalues*

We take H to be either real symmetric, or complex hermitian, or quaternion real self-dual. The matrix H has n real diagonal elements and $\frac{1}{2}n(n-1)$ non-diagonal independent elements which are real, complex or quaternion real. Thus the total number of real variables in H is $n + \frac{1}{2}n(n-1)\beta$. Also H has n real eigenvalues. So if we want the eigenvalues of H as new variables, we have to supplement them with $\frac{1}{2}n(n-1)\beta$ extra variables. What is the best choice? The matrix which diagonalizes H also has $\frac{1}{2}n(n-1)\beta$ (angular) real parameters to characterize it. Choosing these angular parameters and the n (real) eigenvalues of H as new variables, the exponential part of $P(H)$ is easily expressed since $\operatorname{tr} H^j = \sum_{k=1}^{n} \theta_k^j$. One has now to evaluate the Jacobian. The result, after some calculation, is

$$J = \prod_{j<k} |\theta_j - \theta_k|^\beta \cdot f, \tag{11}$$

where θ_j are the eigenvalues of H, β is 1, 2 or 4 for real, complex or quaternion real H, and f depends only on the angular parameters specifying the matrix U which digonalizes H. For anti-symmetric, anti-hermitian or anti-self-dual matrices one should note that the non-zero eigenvalues come in pairs. Changing the origin and the scale, we can thus write the joint probability density of the eigenvalues as

$$P(H) = C(\beta, n) \exp\left(-\sum_{j=1}^{n} x_j^2\right) \prod_{1 \le j < k \le n} |x_j - x_k|^\beta, \tag{12}$$

3. Some general questions which are well posed

The *m-point correlation function* $R_m(x_1,\dots,x_m)$ is defined as

$$R_m(x_1,\dots,x_m)=\frac{n!}{(n-m)!}\int\cdots\int P(x_1,\dots,x_n)\mathrm{d}x_{m+1}\cdots\mathrm{d}x_n. \tag{13}$$

In fact $R_m(x_1,\dots,x_m)\mathrm{d}x_1\cdots\mathrm{d}x_m$ is the probability of having one eigenvalue in each of the m intervals $\mathrm{d}x_1,\dots,\mathrm{d}x_m$ around the points $x_1,\dots,x_m$ respectively. When the m points separate into two or more clusters far away from each other, then R_m is usually a product of functions depending separately on each cluster. To measure the effect of neighborhood on the probabilities one defines cluster functions $T_m(x_1,\dots,x_m)$. They are also known under the name of cumulants and correspond to connected graphs in the diagrammatical representation of the perturbation series due to Ursell, Yvon-Mayer in statistical mechanics, and more spectacularly due to Feynman in field theory. They are zero whenever the m points divide into two or more sets far away from each other. The functions T_m can be expressed in terms of the R_m and conversely. So it is sufficient to evaluate one of these two sets.

Level spacing functions. One might ask for the probability $E(m,t)$ of having exactly m eigenvalues inside an interval of length $s=2t$ chosen at random. Usually the length of the interval is measured in terms of the local mean spacing. Or, one might ask for the probability $F(m,s)$ of having m eigenvalues inside an interval of length s measured from a randomly chosen eigenvalue; or for the probability $p(m,s)\mathrm{d}s$ that the distance between two eigenvalues lies between s and $s+\mathrm{d}s$, while this interval s contains exactly m eigenvalues inside it. These three sets of functions satisfy the relations

$$F(m,s)=-\frac{\mathrm{d}}{\mathrm{d}s}\sum_{j=0}^{m}E(j,t),\quad p(m,s)=-\frac{\mathrm{d}}{\mathrm{d}s}\sum_{j=0}^{m}F(j,s),\quad \text{for } m\geq 0. \tag{14}$$

It will therefore be sufficient to know $E(m,t)$ from which the other functions can be deduced. One may of course ask other questions; for example, what is the average number of eigenvalues in an interval s chosen at random? and what is its dispersion around this average? Or how well a linear graph represents the number of eigenvalues in an interval as a function of its length?

Correlation functions for the Gaussian ensembles. The case $\beta=2$ of complex hermitian random matrices is mathematically the simplest, since we need only real symmetric matrices and their determinants. For the cases $\beta=1$ (most important for applications) and $\beta=4$, one needs self-dual quaternion matrices and their determinants. To evaluate the m-point correlation function we will use the following theorem.

Theorem 3. *Let the complex function* $f(x,y)$ *be such that*

$$\text{(i)}\quad f(x,y)=[f(y,x)]^{*}, \tag{15}$$

$$\text{(iii)} \qquad \int f(x,y)\, f(y,z)\mathrm{d}\mu(y) = f(x,z), \tag{17}$$

then the integral of the $n \times n$ *determinant* $\det[f(x_i,x_j]$ *over* $\mathrm{d}\mu(x_n)$ *is, apart from a constant, equal to the* $(n-1)\times(n-1)$ *determinant obtained from the original one by removing the row and column containing* x_n*; i.e.*

$$\int \det[f(x_i,x_j]_{i,j=1,\ldots,n} = (c-n+1)\det[f(x_i,x_j)]_{i,j=1,\ldots,n-1}. \tag{18}$$

The same theorem is valid when $f(x,y)$ is a quaternion real function satisfying the requirements,

$$\text{(i)} \qquad f(x,y) = [f(y,x)]^{D}, \quad \text{(ii)} \quad \int f(x,x)\mathrm{d}\mu(x) = c, \tag{19}$$

$$\text{(iii)} \qquad \int f(x,y)\, f(y,z)\mathrm{d}\mu(y) = f(x,z) + \lambda f(x,z) - f(x,z)\lambda, \tag{20}$$

where λ is a constant quaternion and the determinant of quaternion matrices is given in Definition 1, eq. (7). As we said earlier the determinant of a quaternion matrix is a convenient way of speaking about Pfaffians of an anti-symmetric matrix. For a proof see chapter 8 of MT. To calculate the m-point correlation function all we have to do is to express

$$P_\beta(x_1,\ldots,x_n) = \exp\left(-\sum_{j=1}^{n} x_j^2\right) \prod_{1\le j<k\le n} |x_j - x_k|^\beta \tag{21}$$

as an $n \times n$ determinant $\det[f(x_i,x_j)]$ with the function $f(x,y)$ satisfying the requirements of the above Theorem 3. When $\beta = 2$, our $f(x,y)$ will be a real function and when β is 1 or 4, it will be a quaternion real function.

3.1 *Alternants*

In the last century Vandermonde evaluated the following determinant

$$\det[x_i^{j-1}]_{i,j=1,\ldots,n} = \prod_{1\le i<j\le n} (x_j - x_i). \tag{22}$$

Actually, recalling the elementary properties of a determinant one can replace the powers of x by any polynomial of degree j. Thus

$$\det[C_{j-1}(x_i)]_{i,j=1,\ldots,n} = \text{constant} \prod_{1\le i<j\le n} (x_j - x_i), \tag{23}$$

where the constant on the right hand side is the product of the coefficients of x^j in $C_j(x)$ for $j = 0, 1, \ldots, n-1$. The polynomials $C_j(x)$ being arbitrary, we can choose them as we

$$\det[Q_{j-1}(x_i),\ Q'_{j-1}(x_i)]\ \begin{matrix} i=1,\ldots,n \\ j=1,\ldots,2n \end{matrix} = \text{constant} \prod_{1\le i<j\le n} (x_j - x_i)^4, \tag{24}$$

where the polynomials $Q_j(x)$ are again arbitrary, $Q'_j(x)$ is the derivative of $Q_j(x)$ and the constant is the product of the coefficients of x^j in $Q_j(x)$ for $j = 0, \ldots, 2n-1$. We will choose three sets of polynomials $R_j(x)$, $C_j(x)$ and $Q_j(x)$ corresponding to the three cases $\beta = 1$, $\beta = 2$ and $\beta = 4$ to express $P_\beta(x_1, \ldots, x_n)$, eq. (21), as an $n \times n$ determinant $\det[f(x_i, x_j]$ satisfying the conditions (15–17) or (20). (The letters R, C and Q are chosen here to recall that they correspond to real, complex and quaternion matrices.)

3.2 *Orthogonal polynomials*

Given a non-negative weight function $w(x)$ for $a \le x \le b$, such that $\int_a^b x^j w(x) \mathrm{d}x$ is finite for every integer $j \ge 0$, one can define a series of polynomials $p_j(x)$ such that

$$\int_a^b p_j(x) p_k(x) w(x) \mathrm{d}x = c_j \delta_{jk}, \tag{25}$$

where δ_{jk} is 1 if $j = k$ and 0 if $j \ne k$. The constants c_j are positive and can be chosen to be 1. The polynomials $p_j(x)$ are called orthogonal; they are called orthonormal if $c_j = 1$. Actually, these polynomials result from the orthogonalization of the power series $(1, x, x^2, \ldots)$ with respect to the symmetric scalar product $(x^i, x^j) = \int_a^b x^i \cdot x^j \cdot w(x) \mathrm{d}x$. But this scalar product need not be symmetric and can be chosen otherwise. For the case $\beta = 2$, we find it convenient to choose the polynomials $C_j(x)$ as the familiar orthogonal polynomials with weight $w(x) = \exp(-x^2)$ for $-\infty < x < \infty$. For the other two cases $\beta = 1$ and $\beta = 4$, we will choose the polynomials $R_j(x)$ and $Q_j(x)$ as skew-orthogonal polynomials resulting from anti-symmetric scalar products defined later. We choose the polynomials $C_j(x)$ to be proportional to the familiar Hermite polynomials $H_j(x) = \mathrm{e}^{x^2}(-\mathrm{d}/\mathrm{d}x)^j \mathrm{e}^{-x^2}$, and the functions

$$\phi_j(x) = (2^j j! \sqrt{\pi})^{-1/2} H_j(x) \exp(-x^2/2), \tag{26}$$

so that $\phi_j(x)$ are orthonormal on $(-\infty, \infty)$,

$$\int_{-\infty}^{\infty} \phi_j(x) \phi_k(x) \mathrm{d}x = \delta_{jk}. \tag{27}$$

Then from eq. (23)

$$\mathrm{e}^{-\frac{1}{2}(x_1^2+\cdots+x_n^2)} \prod_{1\le j<i\le n} (x_i - x_j) = \text{constant} \det[\phi_{j-1}(x_i)]_{i,j=1,\ldots,n}, \tag{28}$$

or, for the case $\beta = 2$,

$$P_2(x_1, \ldots, x_n) = \text{constant} \det[\phi_{\ell-1}(x_j)] \cdot \det[\phi_{\ell-1}(x_k)] \tag{29}$$

$$= \text{constant } \det \left[\sum_{\ell=0}^{n-1} \phi_{\ell-1}(x_j)\phi_{\ell-1}(x_k) \right]_{j,k=1,\dots,n} \tag{30}$$

$$= \text{constant } \det[f(x_j, x_k))], \tag{31}$$

where

$$f(x,y) = \sum_{\ell=0}^{n-1} \phi_\ell(x)\phi_\ell(y). \tag{32}$$

Now let us examine whether this $f(x,y)$ satisfies the conditions of Theorem 3. As $f(x,y)$ is real and symmetric, condition (15) holds true. Also due to orthonormality of the functions $\phi_j(x)$, eq. (27), one has $\int_{-\infty}^{\infty} f(x,x)\mathrm{d}x = n$, and

$$\int_{-\infty}^{\infty} f(x,y)f(y,z)\mathrm{d}y = \sum_\ell \sum_m \phi_\ell(x)\phi_m(z) \int_{-\infty}^{\infty} \phi_\ell(y)\phi_m(y)\mathrm{d}y \tag{33}$$

$$= \sum_\ell \phi_\ell(x)\phi_\ell(z) = f(x,z). \tag{34}$$

Thus we have expressed $P_2(x_1,\dots,x_n)$ as an $n \times n$ determinant $[f(x_j,x_k)]$, where $f(x,y)$ satisfies the conditions of Theorem 3. Using Theorem 3 several times one gets the m-point correlation function in case $\beta = 2$. For the undetermined constant in eq. (31) it is sufficient to integrate over all the n variables and check the normalization of $P_2(x_1,\dots,x_n)$. Thus in case $\beta = 2$,

$$R_m(x_1,\dots,x_m) = \det[f(x_i,x_j)]_{i,j=1,\dots,m}. \tag{35}$$

The m-point cluster function $T_m(x_1,\dots,x_m)$ is then

$$T_m(x_1,\dots,x_m) = \sum f(x_1,x_{i_2})f(x_{i_2},x_{i_3})\cdots f(x_{i_m},x_1), \tag{36}$$

where the sum is taken over all $(m-1)!$ permutations of $(2,3,\dots,m)$.

3.3 *Skew-orthogonal polynomials*

Instead of the symmetric scalar product one may consider an anti-symmetric scalar product

$$(f,g)_Q = \int \left(f(x)\,\frac{\mathrm{d}}{\mathrm{d}x} g(x) - g(x)\frac{\mathrm{d}}{\mathrm{d}x}\, f(x) \right) w(x)\mathrm{d}x \tag{37}$$

or

$$(f,g)_R = \iint f(x)g(y)w(x,y)\mathrm{d}x\,\mathrm{d}y \tag{38}$$

with $w(x,y) = -w(y,x)$. Such a scalar product can be used to construct skew-orthogonal polynomials. We will use them to write $P_\beta(x_1,\dots,x_n)$ for $\beta = 4$ and $\beta = 1$ as an $n \times n$ quaternion determinant. Let us first take the simpler case $\beta = 4$, and choose polynomials

$Q_j(x)$ satisfying the skew-orthogonality relations

$$\int_{-\infty}^{\infty} [Q_{2j}(x)Q'_{2k+1}(x) - Q'_{2j}(x)Q_{2k+1}(x)]\mathrm{e}^{-x^2}\,\mathrm{d}x = \delta_{jk}, \tag{39}$$

$$\int_{-\infty}^{\infty} [Q_{2j}(x)Q'_{2k}(x) - Q'_{2j}(x)Q_{2k}(x)]\mathrm{e}^{-x^2}\,\mathrm{d}x = 0, \tag{40}$$

$$\int_{-\infty}^{\infty} [Q_{2j+1}(x)Q'_{2k+1}(x) - Q'_{2j+1}(x)Q_{2k+1}(x)]\mathrm{e}^{-x^2}\,\mathrm{d}x = 0, \tag{41}$$

i.e.

$$\int_{-\infty}^{\infty} \phi_j^D(x)\phi_k(x)\mathrm{e}^{-x^2}\,\mathrm{d}x = \delta_{jk}, \tag{42}$$

where $\phi_j(x)$ is the quaternion having the 2×2 matrix representation

$$C(\phi_j(x)) = \begin{bmatrix} Q_{2j}(x) & Q_{2j+1}(x) \\ Q'_{2j}(x) & Q'_{2j+1}(x) \end{bmatrix}, \tag{43}$$

and $\phi_j^D(x)$ is its dual. Now choose the quaternion $f_4(x,y)$ as

$$f_4(x,y) = \sum_{j=0}^{n-1} \phi_j(x)\phi_j^D(y). \tag{44}$$

Theorem 4. *We claim that*

$$\det[f_4(x_j,x_k)] = \text{constant} \prod_{1\le j<k\le n} (x_j - x_k)^4, \tag{45}$$

$$\int_{-\infty}^{\infty} f_4(x,x)\mathrm{e}^{-x^2}\,\mathrm{d}x = n, \tag{46}$$

and

$$\int_{-\infty}^{\infty} f_4(x,y)f_4(y,z))\mathrm{e}^{-y^2}\,\mathrm{d}y = f_4(x,z). \tag{47}$$

Proof. From the definition of $f_4(x,y)$, eq. (44), and the orthonormality relation, eq. (42), one has

$$\int_{-\infty}^{\infty} f_4(x,y)f_4(y,z)\mathrm{e}^{-y^2}\,\mathrm{d}y = \sum_{\ell}\sum_{m} \phi_\ell(x)\phi_m^D(z)\int_{-\infty}^{\infty} \phi_\ell^D(y)\phi_m(y)\mathrm{e}^{-y^2}\,\mathrm{d}y \tag{48}$$

$$= \sum_{\ell} \phi_\ell(x)\phi_\ell^D(z) = f_4(x,z), \tag{49}$$

which is (47). Also since $\phi_j(x)$ and $\phi_j^D(x)$ commute (!),

$$\int_{-\infty}^{\infty} f_4(x,x)\mathrm{e}^{-x^2}\,\mathrm{d}x = \sum_{j=0}^{n-1} \int_{-\infty}^{\infty} \phi_j(x)\phi_j^D(x)\mathrm{e}^{-x^2}\,\mathrm{d}x \tag{50}$$

$$= \sum_{j=0}^{n-1} \int_{-\infty}^{\infty} \phi_j^D(x)\phi_j(x)\mathrm{e}^{-x^2}\mathrm{d}x = n. \tag{51}$$

To prove (45), observe that $\det[f_4(x_j, x_k)]$ is equal to the Pfaffian of the $2n \times 2n$ anti-symmetric matrix $C(e_1 I)C([f_4(x_j, x_k)])$, which in turn is the square root of $\det C([f_4(x_j, x_k)])$, i.e.

$$\det[f_4(x_j, x_k)] = \det C([\phi_{j-1}(x_k)]). \tag{52}$$

But from (43) and (24), this is just (45). Thus we have expressed $P_4(x_1, \ldots, x_n)$ as an $n \times n$ quaternion determinant

$$P_4(x_1, \ldots, x_n) = \text{constant } \det[f_4(x_j, x_k)] \cdot \mathrm{e}^{-(x_1^2 + \cdots + x_n^2)}, \tag{53}$$

and the function $f_4(x, y)$ satisfies the requirements of Theorem 3. Hence the m-point correlation function can be written as an $m \times m$ quaternion determinant

$$R_m(x_1, \ldots, x_m) = \text{constant } \mathrm{e}^{-(x_1^2 + \cdots + x_m^2)} \cdot \det[f_4(x_j, x_k)]_{j,k=1,\ldots,m}, \tag{54}$$

or as a Pfaffian of a $2m \times 2m$ real anti-symmetric matrix. The constant in the last equation can be fixed by integrating over all the variables, the result should be 1. Next we take the case $\beta = 1$. The treatment is simpler when n is even, $n = 2r$. Let

$$g_j(x) = \int_{-\infty}^{\infty} \epsilon(x - y) R_j(y) \mathrm{e}^{-y^2} \mathrm{d}y, \tag{55}$$

where $R_j(x)$ is a polynomial of degree j, and $\epsilon(x) = (1/2)\mathrm{sign}(x)$, i.e. it is $1/2$ for $x > 0$, $-1/2$ for $x < 0$ and 0 for $x = 0$. Let the quaternion $\psi_j(x)$ have the 2×2 matrix representation

$$C(\psi_j(x)) = \begin{bmatrix} g_{2j}(x) & g_{2j+1}(x) \\ g'_{2j}(x) & g'_{2j+1}(x) \end{bmatrix}, \tag{56}$$

and we choose the polynomials $R_j(x)$ such that

$$\int_{-\infty}^{\infty} \psi_j^D(x)\psi_k(x)\mathrm{d}x = \delta_{jk}. \tag{57}$$

The quaternion $\sum_{j=0}^{r-1} \psi_j(x)\psi_j^D(y)$ having the 2×2 matrix representation

$$G(x, y) = \begin{bmatrix} S(x, y) & I(x, y) \\ D(x, y) & S(y, x) \end{bmatrix} \tag{58}$$

with

$$S(x, y) = \sum_{j=0}^{r-1} [g_{2j}(x)g'_{2j+1}(y) - g'_{2j}(y)g_{2j+1}(x)]$$

$$D(x, y) = \sum_{j=0}^{r-1} [g'_{2j}(x)g'_{2j+1}(y) - g'_{2j}(y)g'_{2j+1}(x)],$$

$$I(x, y) = \sum_{j=0}^{r-1} [g_{2j}(x)g_{2j+1}(y) - g_{2j}(y)g_{2j+1}(x)], \tag{59}$$

$$\int_{-\infty}^{\infty} G(x,y)G(y,z)\mathrm{d}y = G(x,z), \qquad \int_{-\infty}^{\infty} G(x,x)\mathrm{d}x = n, \tag{60}$$

but its determinant is zero, since the matrix $[G(x_j,x_k)]$ is the product of two matrices $C(\psi_\ell(x_j))$ and $C(\psi_\ell^D(x_k))$ of orders $2n \times n$ and $n \times 2n$ respectively (ℓ taking values 0, 1, $\ldots, r-1$, while j and k take values $1,2,\ldots,n$, and $n = 2r$.) This also shows that the rank of the matrix $[G(x_j,x_k)]$ is n and its last n columns $[I(x_j,x_k),\ S(x_k,x_j)]^T$ are linear combinations of its first n columns $[S(x_j,x_k),\ D(x_j,x_k)]^T$. Therefore if we write

$$C(f_1(x,y)) = G(x,y) - \begin{bmatrix} 0 & \epsilon(x-y) \\ 0 & 0 \end{bmatrix}, \tag{61}$$

then the quaternion $f_1(x,y)$ is the dual of $f_1(y,x)$ and with some algebra one verifies that

$$\int_{-\infty}^{\infty} f_1(x,y)f_1(y,z)\mathrm{d}y = f_1(x,z) + \lambda f_1(x,z) - f_1(x,z)\lambda, \tag{62}$$

with $\lambda = \frac{1}{2}e_1$. For this verification one can conveniently use the 2×2 matrix representations of various quaternions and the fact that $g_j'(x) = R_j(x)\mathrm{e}^{-x^2}$. To see that $P_1(x_1,\ldots,x_n)$ is proportional to $\det[f_1(x_j,x_k)]$, one proceeds as follows. The first n columns of $C([f_1(x_j,x_k)])$ are the same as those of $[G(x_j,x_k)]$, while the last n columns of $C([f(x_j,x_k)])$ differ from those of $[G(x_j,x_k)]$ by $[\epsilon(x_j-x_k),\ 0]^T$. Hence

$$\begin{aligned}
\det C([f_1(x_j,x_k)]) &= \det\begin{bmatrix} S(x_j,x_k) & I(x_j,x_k) - \epsilon(x_j-x_k) \\ D(x_j,x_k) & S(x_k,x_j) \end{bmatrix} \\
&= \det\begin{bmatrix} S(x_j,x_k) & -\epsilon(x_j-x_k) \\ D(x_j,x_k) & 0 \end{bmatrix} \\
&= \det[\epsilon(x_j-x_k)]\cdot\det[D(x_j,x_k)].
\end{aligned}$$

Now

$$\begin{aligned}
\det[D(x_j,x_k)] &= \det[(g_{2\ell}'(x_j)g_{2\ell+1}'(x_k) - g_{2\ell}'(x_k)g_{2\ell+1}'(x_j))] \\
&= \det[g_{2\ell}'(x_j), -g_{2\ell+1}'(x_j)]\begin{bmatrix} g_{2\ell+1}'(x_k) \\ g_{2\ell}'(x_k) \end{bmatrix} \\
&= \{\det[g_{\ell-1}'(x_j)]_{\ell,j=1,\ldots,n}\}^2 \\
&= \{\det[R_{\ell-1}(x_j)]_{\ell,j=1,\ldots,n}\}^2\mathrm{e}^{-2(x_1^2+\cdots+x_n^2)}
\end{aligned} \tag{63}$$

and

$$\det[\epsilon(x_j-x_k)] = 1. \tag{64}$$

Hence

$$\det C([f_1(x_j,x_k)]) = \text{constant}\,(P_1(x_1,\ldots,x_n))^2, \tag{65}$$

or

$$P_1(x_1,\ldots,x_n) = \text{constant}\,\det[f_1(x_j,x_k)]. \tag{66}$$

quaternion determinant or the square root of the determinant of an ordinary $2m \times 2m$ matrix. The m-point cluster function is given by (36) where the function $f(x,y)$ is now replaced by either of the quaternion $f_4(x,y)$ or $f_1(x,y)$, the result being a scalar. The m-point cluster function $T_m(x_1,\ldots,x_m)$ is then

$$T_m(x_1,\ldots,x_m) = \sum f(x_1,x_{i_2})\, f(x_{i_2},x_{i_3}) \cdots f(x_{i_m},x_1), \tag{67}$$

where the sum is taken over all $(m-1)!$ permutations of $(2,3,\ldots,m)$.

3.4 1- *and* 2-*point correlation functions*

The one point function $R_1(x)$ gives the density of eigenvalues at x. For $\beta = 2$, it is

$$R_1(x) = \sum_{j=0}^{n-1} \phi_j^2(x) \tag{68}$$

and in the limit of large n it is the "semi-circle",

$$R_1(x) \approx \begin{cases} \frac{1}{\pi}\sqrt{2n - x^2}, & |x| < \sqrt{2n}, \\ 0, & |x| > \sqrt{2n}. \end{cases} \tag{69}$$

For $\beta = 1$ or $\beta = 4$, $R_1(x)$ differs from (68) and in the limit of large n goes to the "semi-circle", eq. (69). Thus the average distance between the eigenvalues near the origin is $\pi/\sqrt{2n}$. The two-point correlation functions are different for the three cases, even in the limit of large n. When $n \to \infty$, $x_j \to 0$ and $x_j\sqrt{2n}/\pi = y_j$ are constants, one has

$$R_2(x) = 1 - s^2(r), \quad \beta = 2, \tag{70}$$

$$= 1 - s^2(r) - \left(\frac{\mathrm{d}}{\mathrm{d}r}s(r)\right)\cdot \int_r^{\infty} s(z)\mathrm{d}z, \quad \beta = 1, \tag{71}$$

$$= 1 - s^2(2r) + \left(\frac{\mathrm{d}}{\mathrm{d}r}s(2r)\right)\cdot \int_0^{r} s(2z)\mathrm{d}z, \quad \beta = 4, \tag{72}$$

where $r = |y_1 - y_2|$ and $s(r) = \sin(\pi r)/(\pi r)$.

4. Circular ensembles

Three other sets of ensembles (i) symmetric unitary random matrices, (ii) unitary random matrices, and (iii) self-dual unitary random matrices, have been extensively studied. They are defined by the following requirements: (i) The probability density $P(A)$ of a symmetric unitary matrix A should remain invariant under the transformation $A \to W^T A W$, W unitary, (case $\beta = 1$), (ii) The probability density $P(A)$ of a self-dual unitary matrix A should remain invariant under the transformation $A \to W^D A W$, W

unitary, (case $\beta = 4$), (iii) The probability density $P(A)$ of a unitary matrix A should remain invariant under the transformation $A \to UAW$, U and W unitary, (case $\beta = 2$). Actually a small variation in A is characterized in the three cases $\beta = 1, \beta = 4$, and $\beta = 2$ respectively by (i) $\mathrm{d}A = W^T \cdot i\mathrm{d}H \cdot W$, $\mathrm{d}H$ real symmetric, (ii) $\mathrm{d}A = W^D \cdot i\mathrm{d}H \cdot W$, $\mathrm{d}H$ quaternion real self-dual, (iii) $\mathrm{d}A = U \cdot i\mathrm{d}H \cdot W$, $\mathrm{d}H$ complex hermitian.

4.1 *Probability density for the eigenvalues*

The eigenvalues of a unitary matrix A are of the form $\mathrm{e}^{i\theta}$, θ real. The matrix A can be diagonalized by a (i) real orthogonal, (ii) quaternion real symplectic, or (iii) complex unitary, matrix respectively in the three cases. For the change of variables from A to its eigenvalues $\mathrm{e}^{i\theta_j}$ one again has to calculate a Jacobian, which in these cases give

$$P_\beta(\theta_1, \ldots, \theta_n) = \text{constant} \prod_{1 \le j < k \le n} |\mathrm{e}^{i\theta_j} - \mathrm{e}^{i\theta_k}|^\beta, \tag{73}$$

where β is 1, 4 or 2.

4.2 *Correlations and cluster functions*

To compute the correlation function

$$R_m(\theta_1, \ldots, \theta_m) = \frac{n!}{(n-m)!} \int_0^{2\pi} P_\beta(\theta_1, \ldots, \theta_n) \mathrm{d}\theta_{m+1} \cdots \mathrm{d}\theta_n \tag{74}$$

one can again use Theorem 3, if one can express $P_\beta(\theta_1, \ldots, \theta_n)$ as an $n \times n$ (ordinary or quaternion) determinant $[f(\theta_j, \theta_k)]$ with $f(\xi, \eta)$ satisfying the conditions of that theorem, namely, (i) $f(\xi, \eta)$ is the complex conjugate or dual of $f(\eta, \xi)$, (ii) $\int_0^{2\pi} f(\xi, \eta) f(\eta, \zeta) \mathrm{d}\eta = f(\xi, \zeta) + \lambda f(\xi, \zeta) - f(\xi, \zeta)\lambda$, λ a constant quaternion, and (iii) $\int_0^{2\pi} f(\xi, \xi) \mathrm{d}\xi = c$, a real or a scalar number. The construction of such a matrix $[f(\theta_j, \theta_k)]$ is similar to and somewhat simpler than in the case of Gaussian ensembles. The results are as follows. The case $\beta = 2$ is again the simplest, one needs only ordinary matrices and determinants, no quaternions.

$$f_2(\xi, \eta) = \frac{1}{2\pi} \sum_p \mathrm{e}^{ip(\xi - \eta)} = S_n(\xi - \eta), \tag{75}$$

$$S_n(\theta) = \frac{\sin(n\theta/2)}{2\pi \sin(\theta/2)} \tag{76}$$

the sum over p in (75) is over the values

$$p = -\tfrac{1}{2}(n-1),\ -\tfrac{1}{2}(n-3), \ldots,\ \tfrac{1}{2}(n-3),\ \tfrac{1}{2}(n-1). \tag{77}$$

For cases $\beta = 4$ and $\beta = 1$ one needs quaternion matrices and their determinants. The results are

$$C(f_4(\xi, \eta)) = \frac{1}{2} \begin{bmatrix} S_{2n}(\xi - \eta) & I_{2n}(\xi - \eta) \\ D_{2n}(\xi - \eta) & S_{2n}(\eta - \xi) \end{bmatrix}$$

$$C(f_1(\xi,\eta)) = \begin{bmatrix} S_n(\xi-\eta) & I_n(\xi-\eta)-\epsilon_n(\xi-\eta) \\ D_n(\xi-\eta) & S_n(\eta-\xi) \end{bmatrix} \tag{78}$$

with $S_n(\theta)$ given by (76) and

$$D_n(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} S_n(\theta) \tag{79}$$

$$I_n(\theta) = \int_0^\theta S_n(\xi)\mathrm{d}\xi$$

$$\epsilon_n(\theta) = \frac{1}{2\pi}\sum_q (iq)^{-1}\mathrm{e}^{iq\theta}, \tag{80}$$

where the sum over q is over $\pm(n+\frac{1}{2})$, $\pm(n+\frac{3}{2}), \ldots$. When $n \to \infty$, $\theta_j \to 0$, and $n\theta_j/(2\pi) = y_j$ finite, then $S_n(\theta) \to s(r)$, $D_n(\theta) \to \mathrm{d}/\mathrm{d}r\, s(r)$, $I_n(\theta) \to \int_0^r s(\xi)\mathrm{d}\xi$, and the correlation functions $R_m(y_1, \ldots, y_m)$ become identical to those for the corresponding Gaussian ensembles. In other words, the statistical properties of any finite number m of the eigenvalues in the limit of large n are identical for (i) Gaussian orthogonal and circular orthogonal ensembles, (ii) Gaussian unitary and circular unitary ensembles, (iii) Gaussian symplectic and circular symplectic ensembles.

4.3 *Relation between the circular orthogonal and circular symplectic ensembles*

Take $2n$-dimensional matrices in the circular orthogonal ensemble. The joint probability density for its eigenvalues is $\prod_{1\le j<k\le 2n} |\mathrm{e}^{i\theta_j} - \mathrm{e}^{i\theta_k}|$. Ordering the eigenvalues around the unit circle $\theta_1 \le \theta_2 \le \cdots \le \theta_{2n}$, if we integrate over alternate eigenvalues, say $\theta_1, \theta_3, \ldots$, θ_{2n-1}, then we get constant $\prod_{1\le j<k\le n}(\mathrm{e}^{i\theta_{2j}} - \mathrm{e}^{i\theta_{2k}})^4$, which is proportional to the joint probability density of the eigenvalues of an $n \times n$ self-dual unitary matrix. In other words, the statistical properties of n alternate angles θ_j, where $\mathrm{e}^{i\theta_j}$ are the eigenvalues of a symmetric unitary matrix of order $2n \times 2n$ taken from the orthogonal circular ensemble, are identical to those of the n angles ϕ_j, where $\mathrm{e}^{i\phi_j}$ are the eigenvalues of an $n \times n$ quaternion self-dual unitary matrix taken from the circular symplectic ensemble. (Theorem 10.6.1 of RM.) A similar relation holds between a random superposition of two orthogonal ensembles and the unitary ensemble. More about this later.

5. Spacing functions $E_\beta(r,t)$

Among other quantities in the random matrix theory the spacing functions have some importance. The spacing function $E_\beta(r,t)$ is the probability that a randomly chosen interval of length $2t$ contains exactly r eigenvalues inside it. We will be interested in the limit when n is large. The case $\beta = 2$ is again mathematically the simplest. To find $E_2(0,t)$, say, one has to integrate $P_2(x_1, \ldots, x_n)$ over all the variables $x_1, \ldots, x_n$ outside the interval of length 2θ. For simplicity we will chose the interval as $(-\theta, \theta)$:

$$E_2(0,\theta) = \int_{\mathrm{out}} P_2(x_1, \ldots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n. \tag{81}$$

But

$$P_2(x_1,\ldots,x_n) = \frac{1}{n!}(\det[\phi_{j-1}(x_\ell)]^2 \\ = \det M, \tag{82}$$

with

$$M_{jk} = \sum_{\ell=1}^{n} \phi_j(x_\ell)\phi_k(x_\ell). \tag{83}$$

In order to compute $E_2(0,t)$, write the first row of the matrix M as the sum of n terms and the determinant as a sum of n determinants. Since all the other rows of M are symmetric in x_j, all these n determinants will give the same answer on integration. So we can replace the first row of M by $\phi_0(x_1)\phi_k(x_1)$ and multiply the result by n. Now subtracting suitable multiples of the first row from other rows we can eliminate x_1 from all of them. The $n-1$ rows of M now contain $x_2, \ldots, x_n$ symmetrically, therefore we can replace the second row of M by $\phi_1(x_2)\phi_k(x_2)$, multiply the result by $n-1$ and eliminate x_2 from the remaining rows. And so on. Thus we get

$$E_2(0,\theta) = \int_{\text{out}} \det\,[\phi_{j-1}(x_j)\phi_{k-1}(x_j)]_{j,k=1,\ldots,n}\mathrm{d}x_1\cdots\mathrm{d}x_n. \tag{84}$$

Since each variable occurs in only one row, we can integrate over them separately,

$$\begin{aligned} E_2(0,\theta) &= \int_{\text{out}} \det\,[\phi_{j-1}(x_j)\phi_{k-1}(x_j)]\mathrm{d}x_1\cdots\mathrm{d}x_n \\ &= \det\left[\delta_{jk} - \int_{-\theta}^{\theta}\phi_j(x)\phi_k(x)\mathrm{d}x\right]_{j,k=0,\ldots,n-1} \\ &= \prod_{j=0}^{n-1}(1-\lambda_j), \end{aligned} \tag{85}$$

where λ_j are the eigenvalues of the symmetric matrix

$$G_{jk} = \int_{-\theta}^{\theta}\phi_j(x)\phi_k(x)\mathrm{d}x \tag{86}$$

i.e. they are solutions of the algebraic equation,

$$\det[\lambda\delta_{jk} - G_{jk}] = 0, \tag{87}$$

or of the integral equation

$$\lambda\psi(x) = \int_{-\theta}^{\theta} K_n(x,y)\psi(y)\mathrm{d}y, \tag{88}$$

with the kernel

$$K_n(x,y) = \sum_{j=0}^{n-1}\phi_j(x)\phi_j(y). \tag{89}$$

(Verify this! Hint: solutions of (88) are of the form $\sum_{i=0}^{n-1} c_i\phi_i(x)$, with the c_i not all zero.) When $n \to \infty$, $\theta \to 0$, while $\theta\sqrt{2n}/\pi = t$ is finite,

$$\frac{\sqrt{2n}}{\pi} K_n\left(\frac{\pi}{\sqrt{2n}}x, \frac{\pi}{\sqrt{2n}}y\right) \to \frac{\sin \pi(x-y)}{\pi(x-y)}. \tag{90}$$

The average distance between the eigenvalues being $(\pi/\sqrt{2n})$, we measure distances in this unit. The integral (88) in the limit $n \to \infty$ is

$$\lambda\psi(x) = \int_{-t}^{t} K(x,y)\psi(y)\mathrm{d}y, \tag{91}$$

with

$$K(x,y) = \frac{\sin \pi(x-y)}{\pi(x-y)}. \tag{92}$$

Much is known about this kernel in the literature. It is the square of another kernel $(2t)^{-1/2}\mathrm{e}^{i\pi xy/t}$ over $(-t,t)$; it commutes with the differential operator $(x^2-1)\mathrm{d}^2/\mathrm{d}x^2 + 2x\,\mathrm{d}/\mathrm{d}x + \pi^2x^2$ and hence has common eigenfunctions with it. The solutions of (91) are known as spheroidal functions, have been extensively studied and tabulated. The eigenvalues λ of (91) all lie between 0 and 1, can be ordered as $1 \geq \lambda_0 \geq \lambda_1 \geq \lambda_2 \geq \ldots$, and when so ordered the eigenfunctions are alternately even and odd. The limiting form of (85),

$$E_2(0,t) = \prod_{i=0}^{\infty}(1-\lambda_i) \tag{93}$$

is a fast converging infinite product. In (93) above, we have written t for convenience, though the length of the interval is $2t$. The $\lambda_j = \lambda_j(t)$ depend of course on t. To find $E_2(r,\theta)$, one has to evaluate the integral

$$E_2(r,\theta) = \frac{n!}{r!(n-r)!}\int P_2(x_1,\ldots,x_n)\prod_{i=1}^{r} a(x_i)\prod_{j=r+1}^{n}(1-a(x_j))\mathrm{d}x_1\cdots\mathrm{d}x_n, \tag{94}$$

where $a(x) = 1$ if $|x| < \theta$ and $a(x) = 0$ if $|x| > \theta$. Again $P_2(x_1,\ldots,x_n)$, (82), can be written as a sum of determinants. Whenever any variable x_j occurs in two or more rows, these rows are proportional and the determinant is zero. Only those determinants survive in which each variable occurs in one row only, so we can replace $P_2(x_1,\ldots,x_n)$ in (94) by

$$\frac{1}{n!}\sum \det[\phi_{j-1}(x_{\ell_j})\phi_{k-1}(x_{\ell_j})]_{j,k=1,\ldots,n} \tag{95}$$

the sum being over all permutations $(\ell_1,\ldots,\ell_n)$ of $(1,2,\ldots,n)$. As each variable occurs in one and only one row in any of these determinants, we can integrate over x_1, ..., x_r from $-\theta$ to θ and over $(x_{r+1},\ldots,x_n)$ outside the interval $(-\theta,\theta)$. Then we expand each determinant in the Laplace manner according to the r rows containing initially the variables x_1, ..., x_r. This gives

$$E_2(r,\theta) = \frac{1}{r!}\sum \pm \det G(i;j)\det[I-G](j';i'). \tag{96}$$

The indices $i \equiv (i_1, \ldots, i_r)$ are chosen from $(0, \ldots, n-1)$ as also the indices $j \equiv (j_1, \ldots, j_r)$. The indices $(i_1, \ldots, i_r)$ are not ordered, while the indices $(j_1, \ldots, j_r)$ are ordered, $j_1 < j_2 < \cdots < j_r$. The sum in the above equation is taken over all choices of indices satisfying the above conditions. The cofactor $[I - G](j'; i')$ of order $(n - r) \times (n - r)$ is obtained from $[I - G]$ by omiting the rows (i) and columns (j), its determinant is apart from a sign

$$\det [I - G](j', i') = \pm \det [I - G] \cdot \det [(I - G)^{-1}_{i,j}]. \tag{97}$$

Ordering the indices $i_1, \ldots, i_r$, $0 \le i_1 < \cdots < i_r \le n - 1$, and taking into account (97) we can write (96) as

$$E_2(r, \theta) = \det [I - G] \sum_{(i;j)} \det G(i; j) \det [I - G]^{-1} (i; j). \tag{98}$$

Diagonalizing G as in § 4.1 above, in the limit $n \longrightarrow \infty$, $\theta\sqrt{2n}/\pi = t$ constant, one has

$$E_2(r, t) = \prod_{i=0}^{\infty} (1 - \lambda_i) \sum_{(j)} \frac{\lambda_{j_1}}{1 - \lambda_{j_1}} \cdots \frac{\lambda_{j_r}}{1 - \lambda_{j_r}}, \tag{99}$$

the sum being taken over all integers $j_1, \ldots, j_r$, with $0 \le j_1 < j_2 < \cdots < j_r$. If we introduce an extra variable z and write

$$D(z, t) = \prod_{i=0}^{\infty} (1 - z\lambda_i(t)), \tag{100}$$

then (93) and (99) can be written as

$$E_2(r, t) = \left(-\frac{\partial}{\partial z} \right)^r D(z, t)\big|_{z=1}. \tag{101}$$

5.1 *Spacing function for the circular unitary ensemble*

To compute the spacing functions for circular ensembles is as easy or as difficult as for the Gaussian ensembles. From

$$|e^{i\theta} - e^{i\phi}| = \pm 4i e^{-(\theta+\phi)/2} (e^{i\theta} - e^{i\phi}), \tag{102}$$

one has

$$\prod_{1 \le j < k \le n} |e^{i\theta_j} - e^{i\theta_k}| = \text{constant } e^{-\frac{1}{2}(n-1)(\theta_1 + \cdots + \theta_n)} \prod_{1 \le j < k \le n} (e^{i\theta_j} - e^{i\theta_k})$$

$$= \text{constant } \det[e^{ip\theta_j}], \tag{103}$$

with

$$p = -\tfrac{1}{2}(n-1),\ -\tfrac{1}{2}(n-3), \ldots,\ \tfrac{1}{2}(n-3),\ \tfrac{1}{2}(n-1), \tag{104}$$

$$P_2(\theta_1,\ldots,\theta_n) = \text{constant}\ \det[\mathrm{e}^{ip\theta_j}]\det[\mathrm{e}^{-iq\theta_j}]$$
$$= \text{constant}\ \det\left[\frac{1}{2\pi}\sum_{j=1}^{n}\mathrm{e}^{i(p-q)\theta_j}\right], \tag{105}$$

where p and q take values as in (104). The probability that the interval $(-\alpha,\alpha)$ does not contain any eigenvalue of a random unitary matrix taken from the circular unitary ensemble is therefore

$$\begin{aligned} E_2(0,\alpha) &= \text{constant} \int_{\alpha}^{2\pi-\alpha} \det\left[\frac{1}{2\pi}\sum_{j=1}^{n}\mathrm{e}^{i(p-q)\theta_j}\right]\mathrm{d}\theta_1\cdots\mathrm{d}\theta_n \\ &= \det\left[\frac{1}{2\pi}\int_{\alpha}^{2\pi-\alpha}\mathrm{e}^{i(p-q)\theta}\mathrm{d}\theta\right] \\ &= \det\left[\delta_{pq}-\frac{1}{2\pi}\int_{-\alpha}^{\alpha}\cos(p-q)\theta\,\mathrm{d}\theta\right] \\ &= \det\left[\delta_{pq}-\frac{\sin(p-q)\alpha}{\pi(p-q)}\right], \end{aligned} \tag{106}$$

the constant is fixed from the normalization $E_2(0,0)=1$. Also

$$\sum_q [\delta_{pq}-\cos(p\theta)\cos(q\theta)][\delta_{qr}-\sin(q\theta)\sin(r\theta)] = [\delta_{pr}-\cos(p-r)\theta], \tag{107}$$

so that

$$F = \det\left[\delta_{pq}-\frac{1}{2\pi}\int_{-\alpha}^{\alpha}\cos(p-q)\theta\,\mathrm{d}\theta\right] = F_+F_-, \tag{108}$$

with

$$\begin{aligned} F_+ &= \det\left[\delta_{pq}-\frac{1}{2\pi}\int_{-\alpha}^{\alpha}\cos(p\theta)\cos(q\theta)\mathrm{d}\theta\right], \\ F_- &= \det\left[\delta_{pq}-\frac{1}{2\pi}\int_{-\alpha}^{\alpha}\sin(p\theta)\sin(q\theta)\mathrm{d}\theta\right]. \end{aligned} \tag{109}$$

Note that

$$\frac{1}{2\pi}\int_{-\alpha}^{\alpha}\cos(p\theta)\cos(q\theta)\mathrm{d}\theta = \frac{1}{2}\left[\frac{\sin(p-q)\alpha}{\pi(p-q)}+\frac{\sin(p+q)\alpha}{\pi(p+q)}\right], \tag{110}$$

and

$$\frac{1}{2\pi}\int_{-\alpha}^{\alpha}\sin(p\theta)\sin(q\theta)\mathrm{d}\theta = \frac{1}{2}\left[\frac{\sin(p-q)\alpha}{\pi(p-q)}-\frac{\sin(p+q)\alpha}{\pi(p+q)}\right], \tag{111}$$

are the even and odd parts of $\sin(p-q)\alpha/(\pi(p-q))$. When $n\to\infty$, $\alpha\to 0$ and $n\alpha/(2\pi)=t$ is finite, the limits of the determinants D, D_+ and D_- are the values of $D(z,t)$,

$$F(z,t) = \prod_{i=0} (1 - z\lambda_i),$$

$$F_+(z,t) = \prod_{i=0}^{\infty} (1 - z\lambda_{2i}),$$

$$F_-(z,t) = \prod_{i=0}^{\infty} (1 - z\lambda_{2i+1}), \tag{112}$$

where λ_i are the eigenvalues of the integral (91). λ_{2i} and λ_{2i+1} correspond respectively to its even and odd solutions, i.e. they are the eigenvalues of the integral equation

$$\lambda\psi(x) = \int_{-t}^{t} K_{\pm}(x,y)\psi(y)\mathrm{d}y, \tag{113}$$

where $K_+(x,y)$ and $K_-(x,y)$ are the even and odd parts of $K(x,y)$,

$$\begin{aligned} K_{\pm}(x,y) &= \tfrac{1}{2}[K(x,y) \pm K(-x,y)] \\ &= \frac{1}{2}\left[\frac{\sin\pi(x-y)}{\pi(x-y)} \pm \frac{\sin\pi(x+y)}{\pi(x+y)}\right]. \end{aligned} \tag{114}$$

5.2 *Orthogonal ensemble*, $\beta = 1$

For $\beta = 1$, integration over alternate variables is possible as follows. We order the variables as $x_1 \leq x_2 \leq \cdots \leq x_n$, so that

$$\prod_{k<j} |x_j - x_k| = \prod_{k<j} (x_j - x_k) = \det\,[x_j^{k-1}]_{j,k=1,\ldots,n}\,, \tag{115}$$

or

$$P_1(x_1,\ldots,x_n) = \text{constant}\ \det\,[\phi_{j-1}(x_k)]_{j,k=1,\ldots,n}\,, \tag{116}$$

where $x_1 \leq x_2 \leq \cdots \leq x_n$ and $\phi_j(x)$ are orthonormal functions, (26). Note that (i) In $P_1(x_1,\ldots,x_n)$ we have changed the weights from $\exp(-\sum x_j^2)$ to $\exp(-\sum x_j^2/2)$; this does not matter much since one can change the scale of the x_j; (ii) In the determinant in (116) we can replace $\phi_{2j+1}(x)$ by $\phi'_{2j}(x)$ if n is even, $n = 2m$. This is because $\phi'_j(x)$, $\phi_{j-1}(x)$ and $\phi_{j+1}(x)$ have a linear relation (verify!)

$$\sqrt{2}\phi'_j(x) = \sqrt{j}\phi_{j-1}(x) - \sqrt{j+1}\phi_{j+1}(x). \tag{117}$$

To see this, we replace ϕ_{2m-1} in the last row by a linear combination of ϕ'_{2m-2} and ϕ_{2m-3}, and eliminate ϕ_{2m-3} in this row by adding a suitable multiple of a previous row. Then replace the row ϕ_{2m-3} by a linear combination of ϕ'_{2m-4} and ϕ_{2m-5}, eliminate ϕ_{2m-5} in this row and so on, till we reach the row ϕ_1 which we replace by ϕ'_0.

(iii) When n is odd, $n = 2m + 1$, by a similar procedure we can replace the rows ϕ_{2j} by ϕ'_{2j-1}, except for the first row which remains ϕ_0. Let us come back to our calculation of

$$
\begin{aligned}
E_1(0,\theta) &= \int P_1(x_1,\ldots,x_n)\prod_{j=1}^{n} u(x_j)\mathrm{d}x_j \\
&= \text{constant} \int_{x_1\le x_2\le\cdots\le x_n} \det\left[x_j^{k-1}\right] \prod_{j=1}^{n} u(x_j)\mathrm{e}^{-x_j^2/2}\mathrm{d}x_j, \qquad (118)
\end{aligned}
$$

where $u(x) = 1$, if $|x| > \theta$ and $u(x) = 0$, if $|x| < \theta$. With the ordering of the variables one has to multiply the result in the last line by $n!$ which is absorbed in the constant. As x_1 occurs only in one column, we can integrate over it. This replaces the first column by

$$
F_j(x_2) = \int_{-\infty}^{x_2} u(x_1)x_1^j\mathrm{e}^{-x_1^2/2}\mathrm{d}x_1, \quad j = 0,1,\ldots,n-1. \qquad (119)
$$

Now x_2 appears in two columns and so we integrate over x_3 replacing the third column by $F_j(x_4) - F_j(x_2)$. But we already have a column $F_j(x_2)$, add it to the third column, which becomes $F_j(x_4)$. In the same way, integration over x_5, x_7, … replaces the corresponding columns by $F_j(x_6)$, $F_j(x_8)$, …. If n is even, $n = 2m$, each of the remaining variables $x_2, x_4,\ldots,x_{2m}$ occur in two columns, $[F_j(x_{2k}), x_{2k}^j]$. If n is odd, $n = 2m+1$, there is an extra column $F_j(\infty)$ on the extreme right. We have still to integrate over x_2, x_4, …, x_{2m} with the restrictions $x_2 \le x_4 \le \cdots \le x_{2m}$, and each variable occurs in two columns. But interchanging any two variables we interchange two columns with two other columns and the determinant does not change. So the integrand is symmetric in the remaining variables and we can remove their ordering provided we divide the result by $m!$, which will be absorbed in the constant outside.

$$
E_1(0,\theta) = \text{constant} \int_{-\infty}^{\infty} \det[F_j(x_{2k}), u(x_{2k})x_{2k}^j\mathrm{e}^{-x_{2k}^2}]\mathrm{d}x_2\,\mathrm{d}x_4\cdots\mathrm{d}x_{2m}. \qquad (120)
$$

In case n is odd, $n = 2m+1$, there is an extra column $F_j(\infty)$. For simplicity we will take n to be even, $n = 2m$. If we expand the determinant in (120) by the first two columns containing x_2, then the cofactor by the two columns containing x_4, and so on, and integrate over $x_2, x_4,\ldots$, we observe that (1) the integral is a sum of terms, each term being the product of m factors of the form

$$
a_{jk} = \int_{-\infty}^{\infty} [F_j(x)x^k - F_k(x)x^j]u(x)\mathrm{e}^{-x^2/2}\mathrm{d}x = -a_{kj}, \qquad (121)
$$

(2) the indices of the various factors a_{jk} in any term are all distinct, and they are in all $(0,1,\ldots,2m-1)$. (3) We can restrict $j < k$ in each a_{jk}, since if $j > k$, then we replace a_{jk} by $-a_{kj}$. (4) The coefficient of the term $a_{j_1j_2}a_{j_3j_4}\cdots a_{j_{2m-1}j_{2m}}$ is $+1$ or -1 according as the permutation from $(0,1,\ldots,2m-1)$ to $(j_1,j_2,\ldots,j_{2m})$ is even or odd. Therefore the result is a Pfaffian

$$
\begin{aligned}
E_1(0,\theta) &= \text{constant} \sum_{(j)} \pm a_{j_1j_2}a_{j_3j_4}\cdots a_{j_{2m-1}j_{2m}} \\
&= \text{constant Pf } \left[a_{jk}\right]_{j,k=0,1,\ldots,2m-1} \\
&= \text{constant } \left(\det[a_{jk}]_{j,k=0,1,\ldots,2m-1}\right)^{1/2}, \qquad (122)
\end{aligned}
$$

with a_{jk} given by (121). One has from the symmetry

$$a_{2j,2k} = a_{2j+1,2k+1} = 0, \tag{123}$$

so half the matrix elements a_{jk} are zero, and the matrix in (122) has the form

$$A = \begin{bmatrix} 0 & a_{01} & 0 & a_{03} & \dots \\ a_{10} & 0 & a_{12} & 0 & \dots \\ 0 & a_{21} & 0 & a_{22} & \dots \\ a_{30} & 0 & a_{32} & 0 & \dots \\ \cdot & \cdot & \cdot & \cdot & \dots \end{bmatrix}. \tag{124}$$

Collecting the non-zero elements in one corner one has

$$\det[a_{jk}] = (\det[a_{2j,2k+1}])^2. \tag{125}$$

Hence

$$E_1(0,\theta) = \text{constant} \det[a_{2j,2k+1}]_{j,k=0,\dots,m-1}. \tag{126}$$

Without changing anything in the above argument we could have replaced $x_j^{2k}e^{-x_j^2/2}$ in the Vandermonde determinant (118) by $\phi_{2k}(x_j)$ and $x_j^{2k+1}e^{-x_j^2/2}$ by $\phi'_{2k}(x_j)$. With this choice (123) is still valid and

$$a_{2j,2k+1} = \delta_{jk} - \int_{-\theta}^{\theta} \phi_{2j}(x)\phi_{2k}(x)\mathrm{d}x. \tag{127}$$

(Verify it!) Therefore

$$\begin{aligned} E_1(0,\theta) &= \det\left[\delta_{jk} - \int_{-\theta}^{\theta} \phi_{2j}(x)\phi_{2k}(x)\mathrm{d}x\right]_{j,k=0,1,\dots,m-1} \\ &= \prod_{i=0}^{m-1}(1-\lambda_{2i}), \end{aligned} \tag{128}$$

where λ_{2i} are the eigenvalues of the integral equation

$$\lambda f(x) = \int_{-\theta}^{\theta} K_{n+}(x,y)\, f(y)\,\mathrm{d}y, \tag{129}$$

with $K_{n+}(x,y)$ the even part of $K_n(x,y)$,

$$\begin{aligned} K_{n+}(x,y) &= \sum_{j=0}^{m-1} \phi_{2j}(x)\phi_{2j}(y) \\ &= \tfrac{1}{2}[K_n(x,y) + K_n(-x,y)], \end{aligned} \tag{130}$$

with $K_n(x,y)$ given by (89). As $n \to \infty$, while $x\sqrt{2n}/\pi$ and $y\sqrt{2n}/\pi$ are kept finite, the integral (129) goes over to

$$\lambda f(x) = \int_{-t}^{t} K_+(x,y)\, f(y)\,\mathrm{d}y, \tag{131}$$

with

$$K_+(x,y) = \frac{1}{2}\left[\frac{\sin\pi(x-y)}{\pi(x-y)} + \frac{\sin\pi(x+y)}{\pi(x+y)}\right], \tag{132}$$

and we have the result

$$E_1(0,t) = F_+(1,t) \tag{133}$$

with $F_+(z,t)$ given by (112).

The calculations of $E_1(2r,\theta)$ and of $E_1(2r+1,\theta)$ are long and I will not write them here again, referring you to either the relevant pages of RM or to the article by Mehta and des Cloizeaux (see the list of references). The result is

$$E_1(2r,t) = E_1(0,t)\sum_{i_1<\cdots<i_r}\frac{\lambda_{2i_1}}{1-\lambda_{2i_1}}\cdots\frac{\lambda_{2i_r}}{1-\lambda_{2i_r}}\left(1-\sum_{j=1}^{r} f_{2i_j}(t)\int_{-t}^{t} f_{2i_j}(x)\,\mathrm{d}x\right) \tag{134}$$

$$E_1(2r-1,t) = E_1(0,t)\sum_{i_1<\cdots<i_r}\frac{\lambda_{2i_1}}{1-\lambda_{2i_1}}\cdots\frac{\lambda_{2i_r}}{1-\lambda_{2i_r}}\sum_{j=1}^{r} f_{2i_j}(t)\int_{-t}^{t} f_{2i_j}(x)\,\mathrm{d}x, \tag{135}$$

where $f_j(x)$ are the solutions of the integral equation, (91) normalized as

$$\int_{-t}^{t} f_j(x)\, f_k(x)\,\mathrm{d}x = \delta_{j,k}\,. \tag{136}$$

From (134) and (135) one sees that

$$\begin{aligned} E_1(2r,t) + E_1(2r-1,t) &= E_1(0,t)\sum_{i_1<\cdots<i_r}\frac{\lambda_{2i_1}}{1-\lambda_{2i_1}}\cdots\frac{\lambda_{2i_r}}{1-\lambda_{2i_r}} \\ &= \left(-\frac{\partial}{\partial z}\right)^r F_+(z,t)\big|_{z=1} \end{aligned} \tag{137}$$

and

$$\begin{aligned} E_1(2r,t) + E_1(2r+1,t) = E_1(0,t)&\sum_{i_1<\cdots<i_r}\frac{\lambda_{2i_1}}{1-\lambda_{2i_1}}\cdots\frac{\lambda_{2i_r}}{1-\lambda_{2i_r}} \\ &\times\left(1-\sum_{j}\frac{\lambda_{2j}}{1-\lambda_{2j}}\, f_{2j}(t)\int_{-t}^{t} f_{2j}(x)\,\mathrm{d}x\right) \end{aligned} \tag{138}$$

which will be shown later to be equal to

$$\left(-\frac{\partial}{\partial z}\right)^r F_-(z,t)\big|_{z=1} \tag{139}$$

5.3 *Circular orthogonal ensemble*

For the circular ensemble, case $\beta = 1$, one can use exactly the same method of integrating over alternate variables. Let us compute the average

$$H \equiv \left\langle \prod_{\text{alt}} u(\theta_j) \prod_{\text{alt}}{}' v(\theta_k) \right\rangle$$

$$= \int_{-\pi}^{\pi} P_1(\theta_1, \ldots, \theta_{2m})\,\mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{2m} \prod_{\text{alt}} u(\theta_j) \prod_{\text{alt}}{}' v(\theta_k), \tag{140}$$

where $u(\theta)$ and $v(\theta)$ are arbitrary functions defined over $(-\pi, \pi)$ and for simplicity we have taken n even, $n = 2m$. The product $\prod_{\text{alt}}$ is taken over a set of alternate points θ_j as they lie on the unit circle, and $\prod'_{\text{alt}}$ is the product over the remaining alternate points. From (103), taking the θ_j to be ordered $-\pi \leq \theta_1 \leq \theta_2 \leq \cdots \leq \theta_{2m} \leq \pi$, one has

$$H = \text{constant} \int_{-\pi}^{\pi} \det[\mathrm{e}^{ip\theta_j}] \prod_{j=1}^{m} u(\theta_{2j-1}) v(\theta_{2j}) \mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{2m}, \tag{141}$$

As in § 5.2 , we integrate over one set of alternate variables, say θ_1, θ_3, …, θ_{2m-1}, remove the ordering of the remaining variables and then integrate over them. For this purpose we define

$$a_{p,q} = \int\int_{-\pi \leq \theta \leq \phi \leq \pi} u(\theta) v(\phi) (\mathrm{e}^{i(p\theta + q\phi)} - \mathrm{e}^{i(p\phi + q\theta)}) \mathrm{d}\theta\, \mathrm{d}\phi, \tag{142}$$

then as in § 5.2, the result is

$$H^2 = \text{constant } \det[a_{p,q}], \quad p, q = -m + \tfrac{1}{2}, -m + \tfrac{3}{2}, \ldots, m - \tfrac{3}{2}, m - \tfrac{1}{2}. \tag{143}$$

Integration in (142) can be done explicitly for simple functions u, v, and factors like $1/p$ can be removed outside. Also we can invert the order of the columns. So setting

$$b_{p,q} = \frac{ip}{4\pi} a_{p,-q}, \tag{144}$$

one has

$$H^2 = \text{constant } \det[b_{p,q}]. \tag{145}$$

If $u(\theta)v(\phi) = u(-\theta)v(-\phi)$, then $b_{-p,-q} = b_{p,q}$, and there are further reductions, now H itself can be written as a determinant,

$$H = \text{constant } \det[F_{p,q}] \tag{146}$$

with

$$F_{p,q} = b_{p,q} + b_{-p,q}$$

$$= \frac{p}{2\pi} \int\int_{-\pi \leq \theta \leq \phi \leq \pi} u(\theta) v(\phi) [\cos(p\theta) \sin(q\phi) - \cos(p\phi) \sin(q\theta)]\, \mathrm{d}\theta \mathrm{d}\phi. \tag{147}$$

If $u(\theta) = v(\theta) = 1$, then $F_{p,q} = \delta_{p,q}$, and the constant in (146) is 1 for correct normalization. If $u(\theta) = v(\theta) = 1$ for $-\pi + \alpha \leq \theta \leq \pi - \alpha$, and $u(\theta) = v(\theta) = 0$ for

$$F_{p,q} = \delta_{p,q} - \frac{1}{\pi}\int_{-\alpha}^{\alpha} \cos(p\theta)\cos(q\theta)\mathrm{d}\theta$$

$$= \delta_{p,q} - \frac{\sin(p-q)\alpha}{\pi(p-q)} - \frac{\sin(p+q)\alpha}{\pi(p+q)}, \tag{148}$$

and $\det[F_{p,q}]$ in the limit of large m with $m\alpha/\pi = t$ becomes $F_+(1,t)$. Next choose $u(\theta) = v(\theta) = 1$ for $-\pi+\alpha \le \theta \le \pi-\alpha$, $u(\theta) = 0$, $v(\theta) = 2$, for $\pi-\alpha < \theta < \pi+\alpha$. This gives

$$F_{p,q} = \delta_{p,q} - \frac{\sin(p-q)\alpha}{\pi(p-q)} + \frac{\sin(p+q)\alpha}{\pi(p+q)}, \tag{149}$$

and $H = \det[F_{p,q}]$ is the probability for the interval $(\pi-\alpha, \pi+\alpha)$ to contain either 0 or 1 eigenvalue, i.e. $H = E_1(0,\alpha) + E_1(1,\alpha)$. (Why $v(\theta) = 2$ and not 1 in $(\pi-\alpha, \pi+\alpha)$?) In the limit of large m and $m\alpha/\pi = t$, this determinant goes over to $F_-(1,t)$.

6. Random matrices in physics

In physics three discrete symmetry operations parity, charge conjugation and time reversal have gained popularity. Here we will be concerned with only the last, the time reversal operation T. From physical considerations time reversal has to be an anti-unitary operator. In the Schrödinger equation $-i\partial\psi/\partial t = H\psi$, H is often real, so the validity of the same equation with $t \to -t$ means changing the sign of i or taking the complex conjugate. As any antiunitary operator, T can be written as $T = KC$, K a unitary operator and C the complex conjugation operator. When the representation of the states is changed by a unitary operator, $\psi \to U\psi$, T transforms to

$$T \to UTU^{-1} = UTU^{\dagger}, \tag{150}$$

or K transforms to

$$K \to UKU^{T}. \tag{151}$$

Operating twice with T should leave the physical system unchanged, i.e.

$$T^2 = \alpha\cdot 1, \quad |\alpha| = 1, \tag{152}$$

where 1 is the unit operator. Or

$$T^2 = KCKC = KK^*CC = KK^* = \alpha 1. \tag{153}$$

But K is unitary,

$$KK^{\dagger} = 1. \tag{154}$$

So

$$K = \alpha K^T = \alpha(\alpha K^T)^T = \alpha^2 K. \tag{155}$$

Therefore

$$\alpha^2 = 1 \quad \text{or} \quad \alpha = \pm 1, \tag{156}$$

so the unitary K is either symmetric or anti-symmetric, i.e.

$$KK^* = 1, \tag{157}$$

or

$$KK^* = -1. \tag{158}$$

These alternatives correspond respectively to integer or half odd integer spin (see Wigner's text on Group Theory). If K is symmetric, then it can always be written as

$$K = UU^T \tag{159}$$

and a transformation $\psi \to U^{-1}\psi$ performed on the states brings K to unity (eq. (151)).

$$K = 1. \tag{160}$$

After one such transformation is found, further transformations $\psi \to R\psi$ are allowed with $RR^T = RR^\dagger = 1$, to keep (160) valid. If the hamiltonian is invariant under time reversal and has integer spin, a representation of the states can be chosen in which it is a real symmetric matrix, and further transformations of the states by real orthogonal matrices is allowed. This corresponds to the Gaussian orthogonal ensemble (GOE). Also if a system is invariant under space rotations (total spin integer or half odd integer), the hamiltonian H commutes with each of the three components of the total angular momentum, in the standard representation of the angular momentum matrices K can be chosen such that it commutes with H, and an H invariant under T is real symmetric. This also corresponds to the Gaussian orthogonal ensemble (GOE). If K is anti-symmetric, then it can be written as $K = UZU^T$, where the only non-zero elements of Z are the 2×2 blocks $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ along the diagonal. A representation of the states can be chosen such that K reduces to Z. Being unitary K is non-singular and the number of rows and columns in K is even. Representing every thing in terms of quaternions a detailed consideration shows that H must be a quaternion real self-dual matrix and further transformations of the states by only symplectic matrices is allowed. This then corresponds to the Gaussian symplectic ensemble (GSE). A system whose hamiltonian is not invariant under the time reversal T corresponds to the Gaussian unitary ensemble (GUE). Till now we talked about the mathematical methods used to calculate quantities like the correlation functions $R_m(x_1, \ldots, x_m)$ or the spacing functions $E(r, t)$. But when experimental data is compared with the theory, one uses only the 2-point function $R_2(x_1, x_2)$, rarely 3 or 4-point functions and instead of $E(r, t)$ one uses $\mathrm{d}^2E(0, t)/\mathrm{d}s^2 = p(s)$, $s = 2t$, giving the probability density of the spacings between nearest neighbors. To make a quantitative judgement of how well two curves fit together is difficult since one does not know the "thickness" of the theoretical curve. So one resorts to various quantities called "statistics" which can be computed from the experimental data alone and whose average and dispersion can be estimated from the theory. For example, the number variance gives the average value and dispersion of the number of eigenvalues in an interval of given length, its skewness and excess, the so called Δ statistic gives the least square best fit of a straight line to the staircase representation of the increase in the number of eigenvalues as a function of energy, covariance of two consecutive spacings, the f-statistic to estimate how well a given eigenvalue fits in the sequence of eigenvalues belonging to a given ensemble, etc. The experimental data examined so far consisted of (1) slow neutron resonance energies in various nuclei such as rare earth and fissionable nuclei, (2) electromagnetic resonance energies in nearly two dimensional microwave cavities, (3) possible quantum energies of a free particle confined to various shaped two dimensional surfaces (quantum billiards). This extensive study is mostly numerical, (4) ultrasonic resonance frequencies of metallic beams, (5) conductance of heterogeneous quenched metallic wire (mesoscopic systems), (6) zeros of the zeta function on the critical line, (7) distribution of trees in a forest, of district head quarters in various states and countries, of centers of bubbles in a froth, ... To appreciate these one should have a look at the various data fitting curves in the literature.

7. Other ensembles

7.1 *Ensemble of matrices without any symmetry*

People have considered ensembles of real, complex or quaternion matrices which are not restricted to be hermitian or unitary. The probability density of each matrix element is still taken to be Gaussian for simplicity. The eigenvalues of such matrices are complex in general and the question is again what can you say about a few of its eigenvalues. As expected, the case of complex matrices is mathematically the simplest. One takes

$$P(H) = \exp\left[-\sum_{j,k} |H_{jk}|^2\right] = \exp[-\mathrm{tr}\,(H^\dagger H)], \tag{161}$$

$$\mathrm{d}H = \prod_{j,k} \mathrm{d}H_{jk}^{(0)} \mathrm{d}H_{jk}^{(1)}, \tag{162}$$

where $H_{jk} = H_{jk}^{(0)} + iH_{jk}^{(1)}$, and $H_{jk}^{(0)}$ and $H_{jk}^{(1)}$ are real. To say anything about the eigenvalues one has first to change variables from the matrix elements to the eigenvalues z_j and some extra variables p_j. Now $\mathrm{tr}(H^\dagger H)$ contains not only the eigenvalues z_j but the extra variables as well, and one has to choose them carefully so as to facilitate later integrations. I will not go in the details of these calculations. The final result is

$$P(z_1, \ldots, z_n) = \exp\left[-\sum_{j=1}^{n} |z_j|^2\right] \prod_{1 \le j < k \le n} |z_j - z_k|^2. \tag{163}$$

This joint probability density is much easier to work with than the Gaussian unitary ensemble (GUE) case. The reason is the orthogonality of the powers of $z = x + iy$,

$$\int_{-\infty}^{\infty} \exp^{-|z|^2} z^j z^{*k} \mathrm{d}x\,\mathrm{d}y = \pi\, j\,!\delta_{jk}, \tag{164}$$

and one does not have to construct orthogonal polynomials. The computation of correlation functions $R_m(z_1, \ldots, z_m)$ follows the same lines as for the GUE. The next difficult case is that of quaternion matrices. There also the joint probability density for the complex eigenvalues, which now come in complex conjugate pairs, turns out to be

$$P(z_1,\dots,z_n) = \exp\left[-\sum_{j=1}^{n} |z_j|^2\right] \prod_j |z_j - z_j^*|^2 \prod_{1\le j<k\le n} |z_j - z_k|^2 \prod_{1\le j<k\le n} |z_j - z_k^*|^2, \tag{165}$$

and to compute the correlation function $R_m(z_1,\dots,z_m)$ one can use the same method of expressing the right hand side of (165) as a quaternion determinant $\det[f(z_j,z_k)]$, the function $f(z,\xi)$ satisfying certain conditions. For real matrices, when the eigenvalues are all real, the resulting joint probability density for the eigenvalues is the same as for the (GOE). But the eigenvalues of a real matrix are not necessarily all real, and the joint probability density of the eigenvalues has not yet been worked out.

7.2 *An integral over a unitary group*

The following deceptively simple-looking integral has proved to be very useful in dealing with two matrix models and the intermediate matrix ensembles. The integral can be written as

$$\int \exp[-c\,\mathrm{tr}\,(A-B)^2]\, f(b)\mathrm{d}B = k_1 \int \exp\left[-c\sum_{j=1}^{n}(a_j-b_j)^2\right] \frac{\Delta(b)}{\Delta(a)} f(b)\mathrm{d}b, \tag{166}$$

or

$$\int \exp[-c\,\mathrm{tr}\,(A-UBU^{-1})^2]\mathrm{d}U = k_2 \frac{\det[\exp(-c(a_j-b_k)^2)]_{j,k=1,\dots,n}}{\Delta(a)\Delta(b)}, \tag{167}$$

where c, k_1 and k_2 are constants, A and B are complex hermitian $n\times n$ matrices with eigenvalues a_j and b_j respectively, $f(b)$ is a function depending only on the eigenvalues of B, $\Delta(a) = \prod_{j<k}(a_j - a_k)$, similarly for $\Delta(b)$, the integral in (167) is over the group of $n\times n$ unitary matrices, $\mathrm{d}B = \prod_j \mathrm{d}B_{jj} \prod_{j,k} \mathrm{d}B_{jk}^{(0)} \mathrm{d}B_{jk}^{(1)}$ and $\mathrm{d}b = \mathrm{d}b_1 \cdots \mathrm{d}b_n$. No such integral over the orthogonal or symplectic group is known.

7.3 *Intermediate ensembles*

We saw that for the time reversal invariant systems with integral spin or rotational symmetry, GOE is a good model, whereas when time reversal invariance is not valid GUE is a good model. What can one say if the time reversal invariance is slightly broken? One should consider then an ensemble of real symmetric matrices plus a small non-real hermitian part, an ensemble intermediate between GOE and GUE slightly away from GOE. Similarly, one should consider an ensemble of matrices intermediate between GSE and GUE, slightly away from GSE. Thanks to the integral (167) one can study these cases of intermediate ensembles not only "near" GOE or GSE, but anywhere in between. I will not go into the details of this calculation. The conclusion is that the transition from GOE

8. Matrix models

In these models one needs to evaluate certain integrals of the form

$$Z = \int \exp\left[-\mathrm{tr} \sum_i V_i(M^{(i)}) - \sum_{i<j} c_{ij}\, \mathrm{tr}\,(M^{(i)} M^{(j)})\right] \mathrm{d}M^{(1)} \mathrm{d}M^{(2)} \cdots, \tag{168}$$

where $V_i(x)$ are even polynomials in x, the $n \times n$ matrices $M^{(i)}$ vary over one of the three domains; (i) all real symmetric matrices, (ii) all complex hermitian matrices, and (iii) all quaternion self-dual matrices. These domains are characterized by a parameter β taking values $1, 2$ and 4 respectively. The integral Z depends on the size n of the matrices, on the parameters entering the polynomials $V_i(x)$ and the coupling coefficients c_{ij}. Or one may interest oneself in the statistical properties of a few of the eigenvalues of these matrices, the so called correlation functions and spacing functions. Of particular interest is the limit when n is large. Analytical solutions are known for one matrix in the three domains, and for two matrices when the matrices are complex hermitian.

8.1 *One-matrix case*

Consider the partition function

$$Z_\beta(g, n) = \int \exp[-\,\mathrm{tr}\, V(M)] \mathrm{d}M, \tag{169}$$

with $V(x)$ an even polynomial of degree 2ν. The case $V(x) = x^2$ has been extensively studied in the random matrix theory. We take $V(x) = x^2 + gx^4/n$. The case of a general polynomial $V(x)$ is similar though more cumbersome. For $\beta = 1$, we will take M to be a $2n \times 2n$ real symmetric matrix and write

$$Z_1(g, n) = \int \exp[-\tfrac{1}{2}\,\mathrm{tr}\, V(M)] \mathrm{d}M. \tag{170}$$

For $\beta = 2$ and $\beta = 4$, we will keep the definition implicit in (169). In the limit of large n we write

$$-\frac{1}{n^2} \ln \frac{Z_\beta(g, n)}{Z_\beta(0, n)} = E^{(\beta)}(g, n) = E_0^{(\beta)}(g) + \frac{1}{n} E_1^{(\beta)}(g) + \frac{1}{n^2} E_2^{(\beta)}(g) + \cdots. \tag{171}$$

The question is to compute $E_j^{(\beta)}(g)$ successively. Another set of questions is related to the evaluation of the m-point correlation function

$$R_m(x_1, \ldots, x_m) = \int \prod_i \mu(x_i) \prod_{j<k} |x_j - x_k|^\beta \mathrm{d}x_{m+1} \mathrm{d}x_{m+2} \ldots, \tag{172}$$

or to the corresponding spacing function when the variables x_j are outside a given interval, for example. As expected, the case $\beta = 2$ is the simplest and was worked out first. The cases $\beta = 1$ and $\beta = 4$ were taken up later as they involve skew-orthogonal polynomials

The first step is to take the eigenvalues of M (and some other auxiliary parameters) as new variables. This gives

$$Z_\beta(g,n) = K \int \exp\left[-\sum_i V(x_i)\right] \prod_{j<k} |x_j - x_k|^\beta \mathrm{d}x_1 \mathrm{d}x_2 \ldots, \tag{173}$$

where the constant K is independent of g, though it depends on n. Now introduce orthogonal polynomials $C_j(x)$, if $\beta = 2$, and skew-orthogonal polynomials $Q_j(x)$ or $R_j(x)$ of appropriate kind if $\beta = 4$ or $\beta = 1$. It is convenient to take all these polynomials to be monic (*monic* implies that the coefficient of the highest power is one).

$$\begin{aligned}
&\langle C_j, C_k\rangle_C = h_j\delta_{jk}, \\
&\langle Q_{2j}, Q_{2k}\rangle_Q = \langle Q_{2j+1}, Q_{2k+1}\rangle_Q = 0, \quad \langle Q_{2j}, Q_{2k+1}\rangle_Q = q_j\delta_{jk}, \\
&\langle R_{2j}, R_{2k}\rangle_R = \langle R_{2j+1}, R_{2k+1}\rangle_R = 0, \quad \langle R_{2j}, R_{2k+1}\rangle_R = r_j\delta_{jk},
\end{aligned} \tag{174}$$

with the definitions of (symmetric or anti-symmetric) scalar products

$$\begin{aligned}
\langle f_1, f_2\rangle_C &= \int_{-\infty}^{\infty} f_1(x) f_2(x) \mathrm{e}^{-V(x)} \mathrm{d}x, \\
\langle f_1, f_2\rangle_Q &= \frac{1}{2}\int_{-\infty}^{\infty} (f_1(x) f_2'(x) - f_1'(x) f_2(x)) \mathrm{e}^{-V(x)} \mathrm{d}x, \\
\langle f_1, f_2\rangle_R &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_1(x) f_2(y) \epsilon(x-y) \mathrm{e}^{-(V(x)+V(y))/2} \mathrm{d}x\, \mathrm{d}y,
\end{aligned} \tag{175}$$

where $\epsilon(x) = \pm 1/2$ according as x is positive or negative. In terms of the constants h_j, q_j and r_j one has

$$\begin{aligned}
Z_1(g,n) &= K.(2n)! \prod_{j=0}^{n-1} r_j, \\
Z_2(g,n) &= K.n! \prod_{j=0}^{n-1} h_j, \\
Z_4(g,n) &= K.2^n.n! \prod_{j=0}^{n-1} q_j.
\end{aligned} \tag{176}$$

One can express the polynomials $R_j(x)$, $C_j(x)$ or $Q_j(x)$ as linear combinations of those from any other set. If one expresses $R_j(x)$ in terms of $C_j(x)$ or $C_j(x)$ in terms of $Q_j(x)$, then such linear combinations contain ν or $\nu + 1$ terms depending on the parity of j, where 2ν is the order of the polynomial $V(x)$. This is so because $C_j(x)$ is orthogonal to any polynomial of degree less than j, $Q_{2j}(x)$ or $R_{2j}(x)$ is skew-orthogonal to any polynomial of degree less than $2j+1$ and $Q_{2j+1}(x)$ or $R_{2j+1}(x)$ is skew-orthogonal to any polynomial of degree less than $2j$.

9.1 *Recurrence relations for h_j, q_j and r_j*

The polynomials $C_j(x)$ satisfy a three term recurrence relation

$$\begin{aligned} xC_j(x) &= C_{j+1}(x) + \rho_j C_{j-1}(x) \\ \rho_j &= h_j/h_{j-1}. \end{aligned} \tag{177}$$

From now on we will take $V(x) = x^2 + gx^4/n$. To get the recurrence relations for ρ_j, consider the integral

$$I = \int_{-\infty}^{\infty} V'(x) C_j(x) C_{j-1}(x) \mathrm{e}^{-V(x)} \mathrm{d}x\,. \tag{178}$$

This can be evaluated in two ways. Express $V'(x)C_{j-1}(x)$ as a linear combination of $C_{j+2}(x)$, $C_{j+1}(x)$, …, $C_{j-4}(x)$ and use orthogonality to get

$$I = 2[1 + (2g/n)(\rho_{j+1} + \rho_j + \rho_{j-1})]h_j\,. \tag{179}$$

And by partial integration we get

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-V(x)} \frac{\mathrm{d}}{\mathrm{d}x}(C_j(x)C_{j-1}(x))\mathrm{d}x = jh_{j-1}\,. \tag{180}$$

Comparing the last two equations

$$\rho_j[1 + (2g/n)(\rho_{j+1} + \rho_j + \rho_{j-1})] = j/2. \tag{181}$$

Equations (177) and (181) give h_j, ρ_j in terms of the initial values $h_0 = \int_{-\infty}^{\infty} \mathrm{e}^{-V(x)}\mathrm{d}x$, $h_1 = \int_{-\infty}^{\infty} x^2 \mathrm{e}^{-V(x)}\mathrm{d}x$, $\rho_0 = 0$, and $\rho_1 = h_1/h_0$. In the limit of large n one can approximate ρ_j by a continuous function, $\rho_j \approx n\rho(j/n)$, so that

$$\rho(x)[1 + 2g\cdot 3\rho(x)] = x/2 \rightarrow \rho(x) = \frac{\sqrt{1+12xg} - 1}{12g}. \tag{182}$$

Similar recurrence relations can be found for q_j and r_j. By partial integration over x in (175), one has

$$\tfrac{1}{2}\langle V'C_j, R_k\rangle_R = \langle C_j, R_k\rangle_C + \langle C_j', R_k\rangle_R\,. \tag{183}$$

Similarly,

$$\tfrac{1}{2}\langle V'Q_j, C_k\rangle_Q = \langle Q_j; C_k\rangle_Q - \langle Q_j', C_k\rangle_C\,. \tag{184}$$

Writing

$$\begin{aligned} R_{2j}(x) &= C_{2j}(x) + a_j C_{2j-2}(x), \\ R_{2j+1}(x) &= C_{2j+1}(x) + b_{j1} C_{2j-1}(x) + bj2C_{2j-3}(x), \end{aligned}$$

$$C_{2j+1}(x) = Q_{2j+1}(x) + \alpha_j Q_{2j-1}(x),$$

$$C_{2j}(x) = Q_{2j}(x) + \beta_{j1} Q_{2j-2}(x) + \beta j 2 Q_{2j-4}(x), \tag{185}$$

one can derive equations relating a_j, b_{j1}, b_{j2}, h_j and r_j, from which eliminating a_j, b_{j1} and b_{j2} one obtains a recurrence relation for the r_j which contain h_j,

$$\frac{r_{j+2}r_{j+1}r_j}{h_{2j+2}h_{2j+1}h_{2j}} \left(\frac{2g}{n}\right)^3 - (j+1)\frac{r_{j+1}r_j}{h_{2j+2}h_{2j}} \left(\frac{2g}{n}\right) - (j+\tfrac{1}{2})\frac{r_j}{h_{2j+1}} + 1 = 0. \tag{186}$$

Similarly from (184) and (185), one can obtain

$$\frac{h_{2j+1}h_{2j}h_{2j-1}}{q_j q_{j-1} q_{j-2}} \left(\frac{2g}{n}\right)^3 - j\frac{h_{2j+1}h_{2j-1}}{q_j q_{j-1}} \left(\frac{2g}{n}\right) - (j+\tfrac{1}{2})\frac{h_{2j}}{q_j} + 1 = 0. \tag{187}$$

Knowing h_j, r_0 and r_1 one finds r_j, or knowing h_j, q_0 and q_1 one finds q_j. For the correlation functions one expresses $\prod_{j<k} |x_j - x_k|^\beta$ as an $n \times n$ determinant $[f(x_j, x_k)]$ (ordinary in case $\beta = 2$, quaternion in case $\beta = 1$ or $\beta = 4$) , where the function $f(x,y)$ satisfies (i) $f(x,y) = \bar{f}(y,x)$, where $\bar{f}$ is the dual of f, (ii) $\int f(x,y) f(y,z) \mathrm{d}y = f(x,z) + \lambda f(x,z) - f(x,z)\lambda$, λ a constant quaternion, (iii) $\int f(x,x)\mathrm{d}x = c$; then one uses the theorem that for such an $f(x,y)$,

$$\int \det[f(x_j,x_k)]_{j,k=1,\ldots,n} \mathrm{d}x_n = (c - n + 1) \det[f(x_j,x_k)]_{j,k=1,\ldots,n-1}. \tag{188}$$

10. Two matrix models

The following two questions can be answered analytically.

1. Calculation of the integral

$$Z = \int F(A,B) \mathrm{d}A\, \mathrm{d}B, \tag{189}$$

where

$$F(A,B) = \text{constant} \exp[-\mathrm{tr}\,(V_1(A) + V_2(B) + cAB)]. \tag{190}$$

$V_1(x)$ and $V_2(x)$ are even polynomials, A and B are $n \times n$ complex hermitian matrices and the integration is from $-\infty$ to ∞ over all the real parameters entering A and B. In particular, how does the integral Z behave as a function of n, c and the coefficients entering $V_1(x)$ and $V_2(x)$, in the limit of large n.

2. Calculation of the correlation functions and spacing functions of the eigenvalues a_j of A and b_j of B. To answer these questions one needs the integral over the group of $n \times n$ unitary matrices U,

$$\int \exp[\mathrm{tr}\,(A - UBU^{-1})^2] \mathrm{d}U = \text{constant} \frac{\det[\mathrm{e}^{-(a_j-b_k)^2}]}{\Delta(a)\Delta(b)}, \tag{191}$$

where $\Delta(a) = \prod_{j<k}(a_j - a_k)$, and similarly for $\Delta(b)$. The integral in (191) again seems deceptively simple. A similar integral over the real orthogonal group (or over the

symmetric (or for quaternion self-dual) matrices A and B. First step is to deduce from (190) the joint probability density for the eigenvalues a_j of A and b_j of B. In view of (191) this is

$$F(a_j;b_j) = K\cdot\exp\left[-\sum_j (V_1(a_j)+V_2(b_j))\right]\Delta(a)\Delta(b)\det[\mathrm{e}^{-ca_jb_k}]_{j,k=1,\dots,n}. \tag{192}$$

To evaluate (189) one introduces bi-orthogonal monic polynomials $P_j(x)$ and $Q_j(x)$ for non-local weight defined by

$$\iint_{-\infty}^{\infty} \exp\left(-V_1(x)-V_2(y)-cxy\right)P_j(x)Q_k(y)\mathrm{d}x\,\mathrm{d}y = h_j\delta_{jk}. \tag{193}$$

One writes $\Delta(a) = \det[P_{j-1}(x_k)]$, $\Delta(b) = \det[Q_{j-1}(b_k)]$, then from the orthogonality (193) we get

$$Z = K\cdot n!\prod_{j=0}^{n-1} h_j. \tag{194}$$

Recurrence relation for the h_j can be deduced from the recurrence relations of the polynomials $P_j(x)$ and $Q_j(x)$. In the limit of large n, the ratio $\rho_j = h_j/h_{j-1}$ can be approximated by a function of a continuous variable j/n. And finally one finds how Z behaves in that limit. To get the correlation functions of the eigenvalues a_j only (not observing any of the b_j) is also simple following the usual method of writing $\int F(a_j;b_j)\mathrm{d}b_1\cdots\mathrm{d}b_n$ as an $n\times n$ determinant $[f(a_j,a_k)]$, and the presence of the matrix B modifies only the weight in the definition of $f(x,y)$. Similarly, the correlation functions of the eigenvalues b_j only are simple to calculate. To get the mixed correlation functions of a_j and b_j is more complicated, since it seems difficult to express $F(a_j;b_j)$ as the determinant of a $2n\times 2n$ matrix with required properties of its matrix elements. One calculates instead the expected value of $\prod_j(1+u(a_j))(1+v(b_j))$, with $u(x)$ and $v(x)$ small, and differentiate functionally with respect to $u(x)$ and $v(x)$ the required number of times. The final result becomes more and more cumbersome as m increases. The mixed spacing functions are also complicated, though can be evaluated in principle.

11. Selberg's integral

Euler's beta integral

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{195}$$

is, of course, well-known. There are at least three generalizations of this, one known since the last century, and two others due to Selberg. From our point of view, the following one is important.

Theorem 5. *Let*

$$\Phi(x)=\prod_{j=1}^{n}x_j^{\alpha-1}(1-x_j)^{\beta-1}|\Delta(x)|^{2\gamma},$$

$$\Delta(x)=\prod_{1\le i<j\le n}(x_i-x_j),\quad n\ge 2,$$

$$=1,\quad n=1,$$

$$\int_0^1 \mathrm{d}x=\int_0^1\cdots\int_0^1 \mathrm{d}x_1\cdots \mathrm{d}x_n \tag{196}$$

then

$$\langle x_1\cdots x_m\rangle \equiv \int_0^1 x_1\cdots x_m\Phi(x)\mathrm{d}x\Big/\!\!\Big/\int_0^1\Phi(x)\mathrm{d}x$$

$$=\prod_{j=1}^{m}\frac{\alpha+(n-j)\gamma}{\alpha+\beta+(2n-j-1)\gamma}, \tag{197}$$

$$I(\alpha,\beta,\gamma,n)\equiv\int_0^1\Phi(x)\mathrm{d}x=\prod_{j=0}^{n-1}\frac{\Gamma(\alpha+j\gamma)\Gamma(\beta+j\gamma)\Gamma(1+\gamma+j\gamma)}{\Gamma(\alpha+\beta+(n+j-1)\gamma)\Gamma(1+\gamma)}. \tag{198}$$

Proof (Aomoto). From the identity

$$\frac{\partial}{\partial x_1}(x_1x_2\cdots x_m\Phi)=\Phi(x)\left[\alpha x_2\cdots x_m-(\beta-1)\frac{x_1x_2\cdots x_m}{1-x_1}\right.$$
$$\left.+2\gamma\sum_{j=2}^{n}\frac{x_1x_2\cdots x_m}{x_1-x_j}\right], \tag{199}$$

we get on integration

$$0=\alpha\langle x_2\cdots x_m\rangle-(\beta-1)\left\langle\frac{x_1\cdots x_m}{1-x_1}\right\rangle+2\gamma\sum_{j=2}^{n}\left\langle\frac{x_1\cdots x_m}{x_1-x_j}\right\rangle. \tag{200}$$

Similarly, the indentity

$$\frac{\partial}{\partial x_1}(x_1^2x_2\cdots x_m\Phi)=\Phi(x)\left[(\alpha+1)x_1x_2\cdots x_m-(\beta-1)\frac{x_1^2x_2\cdots x_m}{1-x_1}\right.$$
$$\left.+2\gamma\sum_{j=2}^{n}\frac{x_1^2x_2\cdots x_m}{x_1-x_j}\right], \tag{201}$$

gives on integration

$$0=(\alpha+1)\langle x_1\cdots x_m\rangle-(\beta-1)\left\langle\frac{x_1^2x_2\cdots x_m}{1-x_1}\right\rangle+2\gamma\sum_{j=2}^{n}\left\langle\frac{x_1^2x_2\cdots x_m}{x_1-x_j}\right\rangle. \tag{202}$$

$$\alpha\langle x_2\cdots x_m\rangle + 2\gamma\sum_{j=2}^{n}\left\langle\frac{x_1\cdots x_m}{x_1-x_j}\right\rangle = (\alpha+1)\langle x_1\cdots x_m\rangle + (\beta-1)\langle x_1\cdots x_m\rangle + 2\gamma\sum_{j=2}^{n}\left\langle\frac{x_1^2x_2\cdots x_m}{x_1-x_j}\right\rangle. \tag{203}$$

Now interchanging x_1 and x_j

$$\left\langle\frac{x_1\cdots x_m}{x_1-x_j}\right\rangle = -\left\langle\frac{x_jx_2\cdots x_m}{x_1-x_j}\right\rangle = \begin{cases} 0, & \text{if} \quad 2\le j\le m,\\ \frac{1}{2}\langle x_2\cdots x_m\rangle & \text{if} \quad m<j\le n. \end{cases} \tag{204}$$

Similarly,

$$\left\langle\frac{x_1^2x_2\cdots x_m}{x_1-x_j}\right\rangle = -\left\langle\frac{x_j^2x_2\cdots x_m}{x_1-x_j}\right\rangle = \begin{cases} \frac{1}{2}\langle x_1\cdots x_m\rangle & \text{if} \quad 2\le j\le m,\\ \langle x_1\cdots x_m\rangle & \text{if} \quad m<j\le n. \end{cases} \tag{205}$$

Substituting (204) and (205) in (203) one gets, upon simplification

$$\langle x_1\cdots x_m\rangle(\alpha+\beta+\gamma(2n-m-1)) = \langle x_2\cdots x_m\rangle(\alpha+\gamma(n-m)), \tag{206}$$

or

$$\langle x_1\cdots x_m\rangle = \frac{\alpha+\gamma(n-m)}{\alpha+\beta+\gamma(2n-m-1)}\langle x_2\cdots x_m\rangle = \prod_{j=1}^{m}\frac{\alpha+\gamma(n-j)}{\alpha+\beta+\gamma(2n-j-1)} \tag{207}$$

which is (197). Setting $m=n$ in (197), one has

$$I(\alpha+1,\beta,\gamma,n) = \prod_{j=1}^{n}\frac{\alpha+\gamma(n-j)}{\alpha+\beta+\gamma(2n-j-1)}I(\alpha,\beta,\gamma,n) = \prod_{j=0}^{n-1}\frac{\alpha+\gamma j}{\alpha+\beta+\gamma(n+j-1)}I(\alpha,\beta,\gamma,n). \tag{208}$$

Thus one can decrease the value of α by one. But changing $x_j \to 1-x_j$, one sees that

$$I(\alpha,\beta,\gamma,n) = I(\beta,\alpha,\gamma,n), \tag{209}$$

so one can also decrease β by one. If α and β are integers, we can express $I(\alpha,\beta,\gamma,n)$ as $I(1,1,\gamma,n)$ multiplied by a known constant. One finally has to evaluate $I(1,1,\gamma,n)$. For this observe that

$$I(1,1,\gamma,n) = \int_0^1 |\Delta(x)|^{2\gamma}\mathrm{d}x = n!\int_{0\le x_1\le x_2\le\cdots\le x_n\le 1}|\Delta(x)|^{2\gamma}\mathrm{d}x. \tag{210}$$

Now change variables as

$$x_n = t, \quad x_j = ty_j, \quad \text{for} \quad j = 1, \ldots, n-1, \tag{211}$$

so that

$$|\Delta(x)|^{2\gamma} = t^{\gamma n(n-1)} \prod_{j=1}^{n-1} (1 - y_j)^{2\gamma} |\Delta(y)|^{2\gamma},$$
$$\mathrm{d}x = t^{n-1} \mathrm{d}t\, \mathrm{d}y, \tag{212}$$

and

$$\begin{aligned} I(1,1,\gamma,n) &= n! \int_0^1 \mathrm{d}t\, t^{\gamma n(n-1)+n-1} \int_{0 \le y_1 \le y_2 \le \cdots \le y_{n-1} \le 1} \mathrm{d}y \prod_{j=1}^{n-1} (1 - y_j)^{2\gamma} |\Delta(y)|^{2\gamma} \\ &= \frac{1}{(\gamma(n-1)+n)} \frac{n!}{(n-1)!} I(1, 2\gamma+1, \gamma, n-1). \end{aligned} \tag{213}$$

Thus we have related $I(1,1,\gamma,n)$ to $I(1,2\gamma+1,\gamma,n-1)$ or to $I(1,1,\gamma,n-1)$ due to (208) and (209). For α, β and γ positive integers, one can therefore express $I(\alpha,\beta,\gamma,n)$ as $I(1,1,\gamma,1)$ multiplied by a known constant. Finally $I(1,1,\gamma,1) = 1$. Collecting all the constants carefully one gets Selberg's integral (198) for α, β and γ positive integers. For other values of α, β and γ one uses analytical continuation.

(Exercise: Using Aomoto's argument, calculate $\langle x_1 \rangle, \langle x_1 x_2 \rangle, \langle x_1^2 \rangle$ where

$$\langle f(x) \rangle = \int_0^\infty f(x) \Phi(x) \mathrm{d}x \Big/ \int_0^\infty \Phi(x) \mathrm{d}x,$$

and

$$\Phi(x) = |\Delta(x)|^{2\gamma} \prod_{j=1}^{n} (x_j^\alpha \mathrm{e}^{-x_j}).$$

11.1 *Some consequences of Selberg's integral*

One can of course change the limits of integration $(0,1)$ to any finite values (a,b) by a linear change of variables, so the weight $x^{\alpha-1}(1-x)^{\beta-1}$ can be changed to the weights for classical orthogonal polynomials on finite intervals such as Jacobi, Gegenbauer, Chebychev or Legendre. Also by a linear change of variables and taking limits one has integrals of the form

$$\int_0^\infty \mathrm{d}x |\Delta(x)|^{2\gamma} \prod_{j=1}^{n} (\mathrm{e}^{-x_j} x_j^\alpha), \tag{214}$$

or

$$\int_{-\infty}^\infty \mathrm{d}x |\Delta(x)|^{2\gamma} \prod_{j=1}^{n} \mathrm{e}^{-x_j^2}, \tag{215}$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2}=f\left(\frac{\mathrm{d}y}{\mathrm{d}t},y,t\right), \tag{216}$$

where f is polynomial in $\mathrm{d}y/\mathrm{d}t$, rational in y, analytic in t and solutions of (216) do not have any movable singularities except possibly for poles. (A singularity is called movable if its position can be changed by changing the initial conditions.) He found that by a change of the dependent or/and independent variable(s), such an equation can always be reduced to one of the 50 equations listed by him. Out of these 50 equations 44 were known before (functions known as trigonometric, hyperbolic, elliptic, Mathieu, Whittaker, ...), and 6 were new. Actually Painlevé found only 3 of them, the other 3 were found by his student Gambier. These 6 equations are now known as Painlevé equations and their solutions as Painlevé transcendents. They are numbered from 1 to 6. In the standard form they read as:

$$\begin{aligned}
&\text{P-1}: && \frac{\mathrm{d}^2y}{\mathrm{d}t^2}=6y^2+t\\
&\text{P-2}: && \frac{\mathrm{d}^2y}{\mathrm{d}t^2}=2y^3+ty+\alpha\\
&\text{P-3}: && \frac{\mathrm{d}^2y}{\mathrm{d}t^2}=\frac{1}{y}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2-\frac{1}{t}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{1}{t}(\alpha y^2+\beta)+\gamma y^3+\frac{\delta}{y}\\
&\text{P-4}: && \frac{\mathrm{d}^2y}{\mathrm{d}t^2}=\frac{1}{2y}\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2+\frac{3}{2}y^3+4ty^2+2(t^2-\alpha)y+\frac{\beta}{y}\\
&\text{P-5}: && \frac{\mathrm{d}^2y}{\mathrm{d}t^2}=\left(\frac{1}{2y}+\frac{1}{y-1}\right)\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2-\frac{1}{t}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{(y-1)^2}{t^2}\left(\alpha y+\frac{\beta}{y}\right)\\
& && \quad+\gamma\frac{y}{t}+\delta\frac{y(y+1)}{y-1}\\
&\text{P-6}: && \frac{\mathrm{d}^2y}{\mathrm{d}t^2}=\frac{1}{2}\left(\frac{1}{y}+\frac{1}{y-1}+\frac{1}{y-t}\right)\left(\frac{\mathrm{d}y}{\mathrm{d}t}\right)^2-\left(\frac{1}{t}+\frac{1}{t-1}+\frac{1}{y-t}\right)\frac{\mathrm{d}y}{\mathrm{d}t}\\
& && \quad+\frac{y(y-1)(y-t)}{t^2(t-1)^2}\left(\alpha+\beta\frac{t}{y^2}+\gamma\frac{t-1}{(y-1)^2}+\delta\frac{t(t-1)}{(y-t)^2}\right)
\end{aligned} \tag{217}$$

where α, β, γ, δ are com-plex constants. Later people studied more complicated second order and simple third order differential equations to see whether they can find some new transcendents, with no success yet. Why are these equations important for us?

Consider the integral equation

$$\lambda f(x)=\int_{-t}^{t}K(x,y)\,f(y)\mathrm{d}y, \tag{218}$$

with the kernel

$$K(x,y) = \frac{\sin \pi(x-y)}{\pi(x-y)}. \tag{219}$$

The solutions of (218) have a definite parity, they are either even or odd, $f(-x) = \pm f(x)$ (Prove!). Even (odd) solutions are also solutions of an integral equation with the kernel $K_+(x,y)$ $(K_-(x,y))$, with

$$K_\pm(x,y) = \tfrac{1}{2}[K(x,y) \pm K(-x,y)]. \tag{220}$$

The even (odd) solutions and the corresponding eigenvalues λ will be denoted by even (odd) indices. The corresponding Fredholm determinants with an extra factor z will be denoted by

$$F_+(z,t) = \prod_{i=0}^{\infty}(1 - z\lambda_{2i}), \quad F_-(z,t) = \prod_{i=0}^{\infty}(1 - z\lambda_{2i+1}), \tag{221}$$

and

$$F(z,t) = F_+(z,t)F_-(z,t) = \prod_{i=0}^{\infty}(1 - z\lambda_i). \tag{222}$$

The functions $F_\pm(z,t)$ and $F(z,t)$, or rather some combinations of them satisfy certain differential equations, non-linear in general. They are in fact Painlevé equations, P5 or P3. Such equations are important in computing the power series expansion or the asymptotic expansion of functions like $E_\beta(r,t)$. Prior to the knowledge of these differential equations such power series or asymptotic series were computed in certain cases with great efforts using all sorts of properties of the spheroidal functions, Toeplitz determinants and inverse scattering theory known in the literature. Now with the differential equations the task is much easier. The Kyoto school of mathematics (Jimbo, Miwa, Sato and Ueno) has studied such equations in a different context and found that $A(t) = -\frac{1}{2}(\mathrm{d}/\mathrm{d}t)\log F(1,t)$ satisfies the equation

$$\left(\frac{t}{2}\frac{\mathrm{d}^2A}{\mathrm{d}t^2} + \frac{\mathrm{d}A}{\mathrm{d}t}\right)^2 + 4\pi t^2\left(\frac{\mathrm{d}A}{\mathrm{d}t}\right)^2 - 2\frac{\mathrm{d}A}{\mathrm{d}t}\left(t\frac{\mathrm{d}A}{\mathrm{d}t} + A\right)^2 = 0, \tag{223}$$

which can be reduced to a standard P5. McCoy *et al* found that the correlation function in the 2-dimensional Ising model satisfies a P5. Now it is suspected that every critical phenomena in physics will involve a Painlevé transcendent in its description. As Conte once said, the critical phenomena are independent of how you prepare the physical system, and so if it is described by a second order differential equation then the singularities of the equation, if any, must be fixed, i.e. its solution must be either a previously known function or a Painlevé transcendent. Though I knew the existence of (223) as McCoy insisted on its importance in random matrix theory over years every time I saw him, I somehow missed the point for at least 10 years. I now realize that almost all correlation or spacing functions satisfy a Painlevé equation, and I hope you will look for one in dealing with any critical phenomena. Let us come back to the Fredholm determinants $F_\pm(z,t)$. They describe completely the spacing functions for the three classes of matrix ensembles Gaussian orthogonal (GOE), Gaussian unitary

(GUE) and Gaussian symplectic (GSE). In fact, if $E_\beta(r,t)$ is the probability that a randomly chosen interval of length $2t$ contains exactly r eigenvalues of a matrix chosen from the GOE ($\beta = 1$), GUE ($\beta = 2$), or GSE ($\beta = 4$), then

$$\frac{1}{r!}\left(-\frac{\partial}{\partial z}\right)^r [F_+(z,t)F_-(z,t)]|_{z=1} = E_2(r,t)$$

$$F_+(1,t) = E_1(0,t)$$

$$\frac{1}{r!}\left(-\frac{\partial}{\partial z}\right)^r F_+(z,t)|_{z=1} = E_1(2r,t) + E_1(2r-1,t), \quad r > 0$$

$$\frac{1}{r!}\left(-\frac{\partial}{\partial z}\right)^r F_-(z,t)|_{z=1} = E_1(2r,t) + E_1(2r+1,t), \quad r \geq 0$$

$$\frac{1}{2.r!}\left(-\frac{\partial}{\partial z}\right)^r [F_+(z,t) + F_-(z,t)]|_{z=1} = E_4(r,t/2). \tag{224}$$

Suppressing the dependence on z for simplicity, let us write

$$A(t) = -\tfrac{1}{2}[\log F_+(z,t) + \log F_-(z,t)],$$

$$B(t) = -\tfrac{1}{2}[\log F_+(z,t) - \log F_-(z,t)], \tag{225}$$

which then satisfy the relations

$$\frac{\mathrm{d}A}{\mathrm{d}t} = 2B^2, \tag{226}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(tA) = z|S(t)|^2, \tag{227}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}(tB) = z\ \mathrm{Re}\,(S(t))^2, \tag{228}$$

$$2\pi t B = z\ \mathrm{Im}\,(S(t))^2, \tag{229}$$

where the auxiliary complex function $S(t) = S(z,t)$ satisfies the differential equation

$$i\,\frac{\mathrm{d}S}{\mathrm{d}t} + \pi S = \frac{z}{2\pi t} S^*(S^2 - S^{*2}) \tag{230}$$

(S^* is the complex conjugate of S), along with the initial condition $S(0) = 1$. Relation (226) was conjectured in 1965 and was proved by Gaudin in 1966, others were discovered in trying to write (223) as

$$(\mathrm{Re}\,S)^2 + (\mathrm{Im}\,S)^2 = |S|^2. \tag{231}$$

We have four real functions A, B, $\mathrm{Re}\,S$ and $\mathrm{Im}\,S$ and five differential relations among them. As a consequence of (226) and (227) we can write

$$A = z|S|^2 - 2tB^2. \tag{232}$$

Eliminating S and B, we get (223) for A, eliminating S and A, we get for B the equation

$$\left(t\,\frac{\mathrm{d}^2B}{\mathrm{d}t^2} + 2\frac{\mathrm{d}B}{\mathrm{d}t} + 2\pi^2 tB\right)^2 = 4B^2\left[\left(t\frac{\mathrm{d}B}{\mathrm{d}t} + B\right)^2 + 4\pi^2t^2B^2\right]. \tag{233}$$

moreover it is the square of another kernel,

$$K(x,y) = \frac{\sin(x-y)\pi}{(x-y)\pi} = \frac{1}{2t}\int_{-t}^{t} \mathrm{d}\xi \exp\left[i\pi(x-y)\xi/\right]t. \tag{234}$$

Equations (223), (230) and (233) can be transformed to Painlevé equations, either to a P5 or to a P3. Let us start with (230). Set $zS^2 = (R+iI)^2$, where R and I are real. Separating real and imaginary parts one has

$$R = \frac{\pi t I'}{\pi^2 t - 2I^2}, \quad I = \frac{\pi t R'}{2R^2 - \pi^2 t}. \tag{235}$$

This pair of equations expresses R in terms of I and I' and I in terms of R and R'. Such pairs are very common in the study of Painlevé equations as we will see. Elimination of I gives

$$R'' = \frac{2R}{2R^2 - \pi^2 t}R'^2 - \frac{2R^2}{2R^2 - \pi^2 t}\frac{R'}{t} + \frac{R}{t}(2R^2 - \pi^2 t). \tag{236}$$

Elimination of R will give the same equation for I. This equation has singular points at $\pm\pi\sqrt{t/2}$ and ∞. To displace them at 0, 1 and ∞, we put $y = (R + \pi\sqrt{t/2})/(R - \pi\sqrt{t/2})$, then y satisfies a Painlevé 5 (verify this!). Similarly $(I + \pi\sqrt{t/2})/(I - \pi\sqrt{t/2})$ satisfies a standard P5. To get rid of factors π, we will set $\tau = \pi t$, $\zeta = 2z/\pi$, and remove factors $2/\pi$ in A and B. Thus (236) becomes, for example,

$$R'' = \frac{R}{R^2 - \tau}R'^2 - \frac{R^2}{R^2 - \tau}\frac{R'}{\tau} + \frac{R}{\tau}(R^2 - \tau). \tag{237}$$

To relate this P5 with a P3, consider the couple of equations

$$f = \frac{\sqrt{\tau}}{R^2 - \tau}\left(R' - \frac{R}{\tau}\right), \quad R = -\frac{\sqrt{\tau}}{f^2 - 1}\left(f' + \frac{3f}{2\tau}\right). \tag{238}$$

Note that f is expressed in terms of R and R', and R is expressed in terms of f and f'. Eliminating f one gets (237), while eliminating R one gets

$$f'' = \frac{f}{f^2 - 1}f'^2 - \frac{f'}{\tau} - \frac{9}{4\tau^2}\frac{f}{f^2 - 1} + f(f^2 - 1). \tag{239}$$

Set $f = \sqrt{u/(u-1)}$, or $u = f^2/(f^2 - 1)$, so that

$$u'' = \frac{3u}{2u(u-1)}u'^2 - \frac{u'}{\tau} + \frac{9u}{2\tau^2}(u-1)^2 - 2u. \tag{240}$$

This is almost a P5 with $\delta = 0$, and becomes a standard P5 on taking τ^2 as the new independent variable. Consider now the couple of equations

$$\frac{u+1}{u-1} = \omega^2 - \omega' - \frac{2\omega}{\tau}, \quad \omega = \frac{2\tau u}{\tau u' - 3u(u-1)}. \tag{241}$$

Eliminating ω, one gets (240), while eliminating u one gets a P3 for ω. As another example we will take (233), which removing the unnecessary factors π can be written as

$$b'' + 4b = \frac{b}{\tau}\sqrt{b'^2 + 4b^2}, \quad b \equiv 4tB. \tag{242}$$

To see that this is a P5, consider the pair of equations

$$b = \frac{\tau\psi}{\psi - 2\tau}, \quad 2\psi = b' + \sqrt{b'^2 + 4b^2}. \tag{243}$$

Eliminating ψ, we get (10.30), and eliminating b we get

$$\psi'' = \left(\frac{1}{\psi} + \frac{1}{\psi - 2\tau}\right)\frac{\psi'^2}{2} - \left(\frac{\psi}{\psi - 2\tau}\right)\frac{\psi'}{\tau} + \frac{\psi}{\tau}(\psi - 2\tau). \tag{244}$$

To put this last equation in the standard form, let

$$\phi = \frac{\psi - 2\tau}{\psi}, \quad \psi = -\frac{2\tau}{\phi - 1}, \tag{245}$$

so that

$$\phi'' = \left(\frac{1}{2\phi} + \frac{1}{\phi - 1}\right)\phi'^2 - \frac{\phi'}{\tau} - \frac{(\phi - 1)^2}{2\tau^2\phi} + 2\phi. \tag{246}$$

This is almost a P5, and will be a P5 by taking τ^2 as the new independent variable. Consider now the pair of equations

$$\phi = \frac{\eta' + \eta^2 - 1}{\eta' + \eta^2 + 1}, \quad \eta = \frac{2\tau\phi}{\tau\phi' - \phi + 1}. \tag{247}$$

Eliminating η one gets (246) for ϕ and eliminating ϕ one gets a P3 for η in the standard form. Similarly, the equation for A can be related either to a P5, as was done by Jimbo, Miwa, Mori and Sato, or to a P3.

Acknowledgements

The author is thankful to R Ramaswamy for editing these notes for publication.

Reference

[1] M L Mehta, *Random matrices* (Academic Press, New York, 1991) (referred to as RM in the text)
[2] M L Mehta, *Matrix theory* (Hindustan Publishing Corporation, Delhi, India) or Les Editions de Physique, Z.I. de Courtaboeuf, B.P. 112, 91944 Les Ulis, France, 1989 (referred to as MT in the text)
[3] "Method of integration over matrix variables: 1–4", *Comm. Math. Phys.* **79**, 327–340 (1981); *Indian J. Pure Appl. Math.* **22**, 531–546 (1991); *J. Phys.* **A14**, 579–586 (1981); *J. Phys.* **11**, 1093–1108 (1991)
[4] M L Mehta and J des Cloizeaux, The probability for several consecutive eigenvalues of a random matrix, *Indian J. Pure Appl. Math.* **3**, 329–351 (1972)

[5] Some papers on Painlevé equations initiated by Jimbo, Miwa, Sato and Ueno of the Kyoto school of mathematics: a good starting point is "Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent" by M Jimbo, T Miwa and Y Mōri, *Physica D* 80–158 (1980). Also the thesis of B Dietz, University of Essen, Germany (1991) and that of A Edelman, Massachusetts Institute of Technology, Cambridge MA, USA are a good reading

[6] K Aomoto, Jacobi polynomials associated with Selberg integrals, *SIAM J. Math. Anal.* **18**, 545–549 (1987)

[7] E P Wigner, *Group Theory* (Academic Press, New York, 1959)

Hölder exponents of irregular signals and local fractional derivatives

KIRAN M KOLWANKAR and ANIL D GANGAL
Department of Physics, University of Pune, Pune 411 007, India

Abstract. It has been recognized recently that fractional calculus is useful for handling scaling structures and processes. We begin this survey by pointing out the relevance of the subject to physical situations. Then the essential definitions and formulae from fractional calculus are summarized and their immediate use in the study of scaling in physical systems is given. This is followed by a brief summary of classical results. The main theme of the review rests on the notion of local fractional derivatives. There is a direct connection between local fractional differentiability properties and the dimensions/local Hölder exponents of nowhere differentiable functions. It is argued that local fractional derivatives provide a powerful tool to analyze the pointwize behaviour of irregular signals and functions.

Keywords. Fractal; multifractal; Hölder exponent; fractional calculus.

PACS Nos 47.53; 47.52; 02.30; 05.40

1. Introduction

In the 1880's, contrary to the notion existing till then that a continuous function must be differentiable at least at some point, examples were constructed [1, 2] to demonstrate that it is possible for continuous functions to be no where differentiable. Weierstrass's construction provided one such early example. For about three decades after the construction of such functions, they were still considered to be rather pathological cases without any practical relevance to the physical world. Perrin was the first to point out their application in a real physical situation, viz., the problem of Brownian motion. He also prophesied [3] the possibility of a simpler way of handling these functions than differentiable ones.

Recent developments in nonlinear and nonequilibrium phenomena suggest that such irregular functions occur much more naturally and abundantly in formulations of physical theories. The work on Brownian motion (see [4]) showed that the graphs of projections of Brownian paths are nowhere differentiable and have dimension $3/2$. A generalization of Brownian motion, called fractional Brownian motion, [3, 4], is known to give rise to graphs having dimension between 1 and 2. It was also observed that typical Feynmann paths [5, 6], like the Brownian paths, are continuous but nowhere differentiable. In fluid systems, passive scalars advected by a turbulent fluid have been shown [7, 8] to have isoscalar surfaces which are highly irregular, in the limit of diffusion constant going to zero. Also there exist situations in which one has to solve a partial differential equation subject to fractal boundary conditions, such as Laplace equation near a fractal conducting surface. For instance, as noted in reference [9] irregular boundaries may appear to be non-differentiable everywhere upto a certain spatial resolution, and/or may exhibit convolutions over many length scales. Nowhere differentiable functions can be used to model these boundaries.

In the study of dynamical systems theory, attractors of some systems are found to be continuous but nowhere differentiable [10]. We will consider one specific example [10, 11] of the dynamical system which gives rise to such an attractor. Consider the following map.

$$\begin{aligned} x_{n+1} &= 2x_n + y_n \mod 1, \\ y_{n+1} &= x_n + y_n \mod 1, \\ z_{n+1} &= \lambda z_n + \cos(2\pi x_n), \end{aligned} \tag{1}$$

where x and y are taken mod 1 and z can be any real number. In order to keep z bounded, λ is chosen between 0 and 1. The equations for x and y are independent of z. The $x - y$ dynamics are chaotic and are unaffected by the value of z. It was shown in [10] that if $\lambda > 2/(3+\sqrt{5})$, the attractor of this map is nowhere differentiable torus. One more example in 2-dim phase space is given in § 5.

All these irregular functions are best characterized locally by a Hölder exponent. We will use the following general definition of the Hölder exponent $h(y)$ which has been used by various authors [12, 13] recently. The exponent $h(y)$ of a function f at y is given by h such that there exists a polynomial $P_n(x)$ of order n, where n is the largest integer smaller than h, that satisfies

$$|f(x) - P_n(x-y)| = O(|x-y|^h), \tag{2}$$

for x in the neighbourhood of y. This definition serves to classify the behavior of the function at y.

Multifractal measures have been object of many investigations [14–21]. This formalism has met with many applications. Its importance also stems from the fact that such measures are natural measures to be used in the analysis of many phenomena [22, 23]. The first mathematical rigorous results concerning multifractals were given in [24]. It may however happen that the object one wants to understand is a function (e.g. a fractal or multifractal signal) rather than a set or a measure. For instance one would like to characterize the velocity field of fully developed turbulence. Studies on fluid turbulence have shown existence of multifractality in velocity fields of a turbulent fluid at low viscosity [14].

In the case of functions having same Hölder exponent h, with $0 < h < 1$, at every point (Weierstrass function (§ 5) is such an example) it is well known [4] that the box dimension of its graph is $2 - h$. On the other hand, there are functions $f(x)$ which do not have constant exponent h at every point but have a range of such exponents. The set of points having same exponents $\{x \mid h(x) = h\}$ for such a function may even constitute a fractal set. In such situations corresponding functions $f(x)$ are multifractal.

In the light of these findings there has been a renewed interest, chiefly among mathematicians, in studying pointwise behaviour of multifractal functions. New tools have been developed to study local behaviour of functions, which will be of practical use. Notable among them is the use of wavelet transforms [25–28] which have been used with some success to this effect. This transform allows one to study the behaviour of functions at different scales and hence it is popularly known as a mathematical microscope. Some well-known functions have been reanalyzed using these techniques [29–32] and shown to possess a spectrum of Hölder exponents, thereby establishing their multifractality.

In this article we review a more direct method, which uses fractional calculus formalism, to characterize local behaviour of functions. The fractional calculus formalism [33–35] allows taking derivatives of fractional order. However, as pointed out in § 2 below, there is a multiplicity of definitions of fractional derivatives. Hence it is important to recognize the appropriate one which is suitable for extracting the scaling behaviour. Recently it was realized [36] that the concept of local fractional derivative, introduced in § 4 below in a more general setting, is ideally suited for this purpose and that there is a direct quantitative connection between the lack of differentiability of a fractal function and the dimension of its graph. Further it provides an alternative way of characterizing the local Hölder exponent. In this formalism, unlike in wavelet transform, Hölder exponent shows up naturally since the magnitude of local fractional derivative (LFD) jumps when its order crosses a critical order. The main theme of this review is centered around this development.

The organization of the paper is as follows. In § 2 we present a brief review of fractional calculus formalism with its recent applications. The next section discusses fractional differentiability using Weyl's definition and certain classical results. In the fourth section we present our definition of LFD in more general setting and its connection to generalized Taylor series. In § 5 we apply this definition to a specific example, viz, Weierstrass' nowhere differentiable function. It is shown that this function, at every point, is locally fractionally differentiable for all orders below $2-s$, but not so for orders between $2-s$ and 1, where s, $1<s<2$, is the box dimension of the graph of the function. In the next section a general result has been proved, showing relation between local fractional differentiability and local Hölder exponent. In § 5 we demonstrate use of LFD in unmasking isolated singularities and in the study of pointwise behavior of multifractal functions. Section 6 concludes the article, pointing out a few possible consequences of our results.

2. Fractional calculus and scaling phenomena

Fractional calculus [33–35] is a study which deals with generalization of differentiation and integration to fractional orders. There are number of ways (not neccesarily equivalent) of defining fractional derivatives and integrations. We list here some of them, which will be used in this work. We begin by recalling the Riemann–Liouville definition of fractional integral of a real function, which is given by [33, 35]

$$\frac{\mathrm{d}^q f(x)}{[\mathrm{d}(x-a)]^q} = \frac{1}{\Gamma(-q)}\int_a^x \frac{f(y)}{(x-y)^{q+1}}\mathrm{d}y \quad \text{for} \quad q<0, \tag{3}$$

where the lower limit a is some real number and of the fractional derivative

$$\frac{\mathrm{d}^q f(x)}{[\mathrm{d}(x-a)]^q} = \frac{1}{\Gamma(n-q)} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \int_a^x \frac{f(y)}{(x-y)^{q-n+1}} \mathrm{d}y \quad \text{for} \quad n-1 < q < n. \tag{4}$$

Fractional derivative of a simple function $f(x) = x^p,\ p > -1$ is given by [33, 35]

$$\frac{\mathrm{d}^q x^p}{\mathrm{d}x^q} = \frac{\Gamma(p+1)}{\Gamma(p-q+1)} x^{p-q} \quad \text{for} \quad p > -1. \tag{5}$$

For future use we note the following properties of fractional derivatives. For $0 < q < 1$,

$$\frac{\mathrm{d}^q \sin(x)}{\mathrm{d}x^q} = \frac{\mathrm{d}^{q-1} \cos(x)}{\mathrm{d}x^{q-1}} \tag{6}$$

and

$$\frac{\mathrm{d}^q \cos(x)}{\mathrm{d}x^q} = \frac{x^{-q}}{\Gamma(1-q)} - \frac{\mathrm{d}^{q-1} \sin(x)}{\mathrm{d}x^{q-1}}. \tag{7}$$

Further the fractional derivative has the property (see ref. [33]), viz,

$$\frac{\mathrm{d}^q f(\beta x)}{\mathrm{d}x^q} = \beta^q \frac{\mathrm{d}^q f(\beta x)}{\mathrm{d}(\beta x)^q} \tag{8}$$

which makes it suitable for the study of scaling. Note the nonlocal character of the fractional derivative and integral in the defining equations (3) and (4). Also it is clear from equation (5) that unlike in the case of integer derivatives the fractional derivative of a constant is not zero in general.

Since the definition of fractional derivative contains a lower limit a, it is natural that different definitions of a fractional differentiability would arise depending on the lower limit one chooses. Weyl defined fractional derivatives, with the arbitrary limit a in (4) going to $-\infty$, as follows.

$$D^q f(x) = \frac{1}{\Gamma(n-q)} \frac{\mathrm{d}^n}{\mathrm{d}x^n} \int_{-\infty}^x \frac{f(y)}{(x-y)^{q-n+1}} \mathrm{d}y \quad \text{for} \quad n-1 < q < n. \tag{9}$$

The merit of this definition is that the fractional derivative of a periodic function according to this definition is a periodic function, which is why it is suitable in harmonic analysis.

2.1 *Applications to the study of scaling in physical systems*

Recent work shows that the fractional calculus formalism is useful in dealing with scaling phenomena. The purpose of this subsection is to point out a few relevant parallel developments even though they are not part of the central theme of this paper.

Hilfer [37–39] has generalized the Eherenfest's classification of phase transition using the fractional derivative. Some recent papers [40–43] indicate a connection between fractional calculus and fractal structures [3, 4] or fractal processes [44–46]. Glöckle and

Fractional Brownian motion: Mandelbrot and Van Ness [44] have used fractional integrals to formulate fractal processes such as fractional Brownian motion. The fractional Brownian motion is described by the probability distribution $B_H(t)$ defined by

$$B_H(t) - B_H(0) = \frac{1}{\Gamma(H+1)} \int_{-\infty}^{t} K(t-t')\mathrm{d}B(t'), \quad 0 < H < 1, \tag{10}$$

where $B(t)$ is an ordinary Gaussian random process with average zero and unit variance and

$$K(t-t') = \begin{cases} (t-t')^{H-1/2} & 0 \le t' < t \\ \{(t-t')^{H-1/2} - (-t)^{H-1/2}\} & t' < 0. \end{cases} \tag{11}$$

The use of the fractional integral kernel is apparent. Recently Sebastian [47] has given a path integral representation for fractional Brownian motion whose measure has fractional derivatives of paths in it.

Fractional equations for a class of Lévy type probability densities: Lévy flights [48] and Lévy walks [49, 50] have found applications in various branches of physics [51], for example in fluid dynamics [52] and polymers [53]. In [40], Nonnenmacher considered a class of normalized one sided Lévy type probability densities, which provides the length of a jump of a random walker, given by

$$f(x) = \frac{a^{\mu}}{\Gamma(\mu)} x^{-\mu-1} \exp(-a/x) \quad a > 0,\ x > 0. \tag{12}$$

It is clear that for large x

$$f(x) \sim x^{-\mu-1}, \quad \mu > 0, \quad x > 0, \tag{13}$$

where μ is the Lévy index. In [40] it was shown that these Lévy-type probability densities satisfy a fractional integral equation

$$x^{2q} f(x) = a^{q} \frac{\mathrm{d}^{-q} f(x)}{\mathrm{d}x^{-q}} \tag{14}$$

or equivalently a fractional differential equation given by

$$\frac{\mathrm{d}^{n-\mu} f(x)}{\mathrm{d}x^{n-\mu}} = a^{-\mu} \frac{\mathrm{d}^{n}}{\mathrm{d}x^{n}} (x^{2\mu} f(x)). \tag{15}$$

Here, $n = 1$ if $0 < \mu < 1$ and $n = 2$ if $1 < \mu < 2$, etc. An interesting observation of this work is that the Lévy index is related to the order of the fractional integral or differential operator.

Fractional equations for physical processes: Modifications of equations governing physical processes such as diffusion equation, wave equation and Fokker–Plank equation have been suggested [41, 42, 54–56] which incorporate fractional derivatives with respect to time. In refs [41, 42] a fractional diffusion equation has been proposed for the diffusion on fractals. Asymptotic solution of this equation coincides with the result obtained numerically. Fogedby [57] has considered a generalization of the Fokker–Plank equation involving addition of a fractional gradient operator (defined as the Fourier transform of $-k^{\mu}$) to the usual Fokker–Plank equation and performed a renormalization group analysis.

Transport in chaotic Hamiltonian systems: Recently Zaslavasky [58] showed that the Hamiltonian chaotic dynamics of particles can be described by a fractional generalization of the Fokker–Plank–Kolmogorov equation which is defined by two fractional critical exponents (α, β) responsible for the space and time derivatives of the distribution function correspondingly. With certain assumptions, the following equation was derived for the transition probabilities $P(x,t)$:

$$\frac{\partial^{\beta} P(x,t)}{\partial t^{\beta}} = \frac{\partial^{\alpha}}{\partial(-x)^{\alpha}}(A(x)P(x,t)) + \frac{1}{2}\frac{\partial^{2\alpha}}{\partial(-x)^{2\alpha}}(B(x)P(x,t)), \tag{16}$$

where

$$A(x) = \lim_{\Delta t \to 0} \frac{A'(x;\Delta t)}{(\Delta t)^{\beta}} \tag{17}$$

and

$$B(x) = \lim_{\Delta t \to 0} \frac{B'(x;\Delta t)}{(\Delta t)^{\beta}} \tag{18}$$

with A' and B' being αth and 2αth moments respectively. The exponents α and β have been related to anomalous transport exponent.

We remark that most of the applications reviewed in this section deal with asymptotic scaling only.

3. Weyl fractional differentiability and Hölder classes

The purpose of this section is to review the classical work of Stein, Zygmund and others [59–62] which uses Weyl fractional calculus in the analysis of irregular functions. Let $f(x)$ be defined in a closed interval I, and let

$$\omega(\delta) = \omega(\delta : f) = \sup |f(x) - f(y)| \quad \text{for} \quad x, y \in I \quad |x - y| \leq \delta. \tag{19}$$

The function $\omega(\delta)$ is called the modulus of continuity of f. If for some $\alpha > 0$ we have $\omega(\delta) \leq C\delta^{\alpha}$ with C independent of δ, then f is said to satisfy Hölder (in old literature it is called Lipschitz) condition of order α. These functions define a class of function Λ_{α}. Notice that with this definition of the modulus of continuity only the case $0 < \alpha \leq 1$ is interesting, since if $\alpha > 1$, $f'(x)$ is zero everywhere and the function is constant. Hence this definition has been generalized (see [28]) to one where the case $\alpha > 1$ is also nontrivial. The idea is to subtract an appropriate polynomial from a function f at y.

$$\tilde{\omega}(\delta) = \tilde{\omega}(\delta : f) = \sup |f(x) - P(x - y)| \quad \text{for} \quad x, y \in I \quad |x - y| \leq \delta, \tag{20}$$

where P is the only polynomial of smallest degree which gives the smallest order of magnitude for $\tilde{\omega}$. Recently, this definition has been widely used [12, 13] to characterize the velocity field of a turbulent fluid.

Following Welland [62] we introduce

DEFINITION 1

f is said to have an αth derivative, where $0 \leq k < \alpha < k+1$ with integer k, if $D^{-\beta}f$, $\beta = k+1-\alpha$, has a $k+1$ Peano derivatives at x_0 i.e. there exists a polynomial $P_{x_0}(t)$ of degree $\leq k+1$ such that

$$(D^{-\beta}f)(x_0+t) - P_{x_0}(t) = o(|t|^{k+1}) \quad t \to 0. \tag{21}$$

Further if

$$\left\{\frac{1}{\rho}\int_{-\rho}^{\rho} |(D^{-\beta}f)(x_0+t) - P_{x_0}(t)|^p \mathrm{d}t\right\}^{1/p} = o(\rho^{k+1}) \quad \rho \to 0 \quad (1 \leq p < \infty), \tag{22}$$

f is said to have an αth derivative in the L^p sense. The $D^{-\beta}$ is in the Weyl sense.

It is clear that this definition of fractional differentiability is not local. Particularly the behavior of function at $-\infty$ also plays a crucial role. The main classical results can be stated using this notion of differentiability and involve the classes Λ_α^p and N_α^p which are given by the following definitions.

DEFINITION 2

If there exists a polynomial $Q_{x_0}(t)$ of degree $\leq k$ s. t. $f(x_0+t) - Q_{x_0}(t) = O(|t|^\alpha)$ as $t \to 0$ then f is said to satisfy the condition Λ_α and if

$$\left\{\frac{1}{\rho}\int_{-\rho}^{\rho} |f(x_0+t) - Q_{x_0}(t)|^p \mathrm{d}t\right\}^{1/p} = O(\rho^\alpha), \quad \rho \to 0 \quad (1 \leq p < \infty), \tag{23}$$

f is said to satisfy the condition Λ_α^p.

DEFINITION 3

f is said to satisfy the condition N_α^p if for some $\rho > 0$

$$\frac{1}{\rho}\int_{-\rho}^{\rho} \frac{|f(x_0+t) - Q_{x_0}(t)|^p}{|t|^{1+p\alpha}} \mathrm{d}t < \infty. \tag{24}$$

We are now in a position to state the classical results in the form of theorems 1 to 4. Next two theorems [61, 62] give the condition under which the fractional derivative of a function exists.

Theorem 1. *Suppose that f satisfies the condition Λ_α at every point of a set E of positive measure. Then $D^\alpha f(x)$ exists almost everywhere in E if and only if f satisfies condition N_α^2 almost everywhere in E.*

The following results in [39] tell us how the class of a function changes when an operation of fractional differentiation is performed.

Theorem 3. *Let* $0 \leq \alpha < 1$, $\beta > 0$ *and suppose that* $f \in \Lambda_\alpha$. *Then* $D^{-\beta} f \in \Lambda_{\alpha+\beta}$ *if* $\alpha + \beta < 1$.

Theorem 4. *Let* $0 < \gamma < \alpha < 1$. *Then* $D^\gamma f \in \Lambda_{\alpha-\gamma}$ *if* $f \in \Lambda_\alpha$.

Though they have their own value, these results are not really adequate to obtain information regarding irregular behaviour of functions and Hölder exponents. We observe that the Weyl definition involves highly nonlocal information and hence is somewhat unsuitable for treatment of local scaling behaviour. In the next section we introduce a more appropriate definition.

4. Local fractional differentiability

In our previous work [36] we introduced the notion of local fractional derivative and demonstrated its use in the study of local scaling behaviour. We now briefly explain this notion and use it in subsequent sections.

4.1 *Local fractional derivative and critical order*

Recall the observations made in § 2, viz, (1) nonlocal information contained in fractional derivative and (2) the fractional derivative of a constant is not zero. The appropriate new notion of differentiability must bypass the hindrance due to these two properties. These difficulties can be remedied by introducing the following.

DEFINITION 4

If, for a function $f : [0, 1] \to \mathbb{R}$, the limit

$$\mathbb{D}^q f(y) = \lim_{x \to y} \frac{\mathrm{d}^q (f(x) - f(y))}{\mathrm{d}(x-y)^q} \tag{25}$$

exists and is finite, then we say that the local fractional derivative (LFD) of order q $(0 < q < 1)$, at $x = y$, exists.

In the above definition the lower limit y is treated as a constant. The subtraction of $f(y)$ corrects for the fact that the fractional derivative of a constant is not zero. Whereas the limit as $x \to y$ is taken to remove the nonlocal content. Advantage of defining local fractional derivative in this manner lies in its local nature and hence allowing the study of pointwise behaviour of functions. This will be seen more clearly in § 4.2 after the development of Taylor series.

DEFINITION 5

We define critical order α, at y, as

$$\alpha(y) = \mathrm{Sup}\{q \mid \text{all LFDs of order less than } q \text{ exist at } y\}.$$

If, for a function $f : [0,1] \to \mathbb{R}$, the limit

$$\mathbb{D}^q f(y) = \lim_{x\to y} \frac{\mathrm{d}^q(f(x) - \sum_{n=0}^{N} (f^{(n)}(y)/\Gamma(n+1))(x-y)^n)}{[\mathrm{d}(x-y)]^q} \tag{26}$$

exists and is finite, where N is the largest integer for which Nth derivative of $f(x)$ at y exists and is finite, then we say that the local fractional derivative (LFD) of order q $(N < q < N+1)$, at $x = y$, exists.

We consider this as the generalization of the local derivative for order greater than one. The definition of the critical order remains the same since, for $q < 1$, (25) and (26) agree. This definition extends the applicability of LFD to functions where the first derivative exists but are still irregular due to the nonexistence of some higher order derivative. As an example we note that the critical order of $f(x) = a + bx + c|x|^\gamma$ at origin, according to definitions 5 and 6 is γ.

Remark. 1) It is interesting to note that the same definition of LFD can be used for negative values of the critical order between -1 and 0. For this range of critical orders $N = -1$ and the sum in (26) is empty. As a result the expression for LFD becomes

$$\mathbb{D}^q f(y) = \lim_{x\to y} \frac{\mathrm{d}^q f(x)}{[\mathrm{d}(x-y)]^q} \,. \tag{27}$$

An equivalence between the critical order and the Hölder exponent, for positive values of critical order, will be proved in § 6. Here we would like to point out that the negative Hölder exponents do arise in real physical situation of turbulent velocity field (see [63, 64] and references therein).

2) Another way of generalizing the LFD to the values of critical order beyond 1 would have been to write it as

$$\mathbb{D}^q f(y) = \lim_{x\to y} \frac{\mathrm{d}^q(f^{(N)}(x) - f^{(N)}(y))}{[\mathrm{d}(x-y)]^q} \,. \tag{28}$$

But the existence of Nth derivative of f at x may not be guaranteed in general. Such a situation may arise in the case of multifractal functions to be treated in § 8.

4.2 *LFD and generalized Taylor series*

In order to see the information contained in the LFD we consider fractional Taylor's series with a remainder term for a real function f. Let

$$F(y, x-y; q, N) = \frac{\mathrm{d}^q(f(x) - \sum_{n=1}^{N} (f^{(n)}(y)/\Gamma(n+1))(x-y)^n)}{[\mathrm{d}(x-y)]^q} \tag{29}$$

and

$$\widetilde{F_N}(x,y) = f(x) - \sum_{n=0}^{N} \frac{f^{(n)}(y)}{\Gamma(n+1)}(x-y)^n. \tag{30}$$

It is clear that

$$\mathbb{D}^q f(y) = F(y,0;q,N). \tag{31}$$

Now, for $N < q < N+1$, it can be shown that

$$\begin{aligned}\widetilde{F_N}(x,y) = {} & \frac{\mathbb{D}^q f(y)}{\Gamma(q+1)}(x-y)^q \\ & + \frac{1}{\Gamma(q+1)} \int_0^{x-y} \frac{\mathrm{d}F(y,t;q,N)}{\mathrm{d}t}(x-y-t)^q \mathrm{d}t,\end{aligned} \tag{32}$$

i.e.

$$f(x) = \sum_{n=1}^{N} \frac{f^{(n)}(y)}{\Gamma(n+1)}(x-y)^n + \frac{\mathbb{D}^q f(y)}{\Gamma(q+1)}(x-y)^q + R_1(x,y), \tag{33}$$

where $R_1(x,y)$ is a remainder given by

$$R_1(x,y) = \frac{1}{\Gamma(q+1)} \int_0^{x-y} \frac{\mathrm{d}F(y,t;q,N)}{\mathrm{d}t}(x-y-t)^q \mathrm{d}t. \tag{34}$$

We note that the local fractional derivative as defined above (not just fractional derivative), along with the first N derivatives, provides an approximation of $f(x)$ in the vicinity of y. We further note that the terms on the RHS of eq. (32) are nontrivial and finite only in the case $q = \alpha$, the critical order. In ref. [65] a fractional Taylor series was constructed by Osler using usual (not local in the present sense) fractional derivatives. His results are, however, applicable to analytic functions and cannot be used for non-smooth functions directly. Further Osler's formulation involves terms with negative orders also and hence is not suitable for approximating schemes.

When $0 < q < 1$ we get as a special case

$$f(x) = f(y) + \frac{\mathbb{D}^q f(y)}{\Gamma(q+1)}(x-y)^q + \text{Remainder} \tag{35}$$

provided the RHS exists. One may note in equation (35) that when q is set equal to one in the above approximation one gets the equation of the tangent. It may be recalled that all the curves passing through a point y and having same tangent form an equivalence class (which is modelled by a linear behavior). Analogously all the functions (curves) with the same critical order α and the same $\mathbb{D}^\alpha$ will form an equivalence class modelled by power law x^α. This is how one may generalize the geometric interpretation of derivatives in terms of tangents. This observation is useful when one wants to approximate an irregular function by a piecewise smooth (scaling) function.

5. Fractional differentiability of Weierstrass function

Consider a form of Weierstrass function, viz,

$$W_\lambda(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin \lambda^k t, \quad \lambda > 1, \quad 1 < s < 2, \quad t \text{ real}. \tag{36}$$

Note that $W_\lambda(0) = 0$. The box dimension of the graph of this function is s. The Hausdorff dimension of its graph is still unknown. The best known bounds are given by Mauldin and Williams [66] where they have shown that there is a constant c such that

$$s - \frac{c}{\log \lambda} \leq \dim_H \text{graph} f \leq s.$$

Recently it was shown [67] that if a random phase is added to each sine term in (36) then the Hausdorff dimension of the graph of the resulting function is s. These kind of functions have been studied in detail in [4, 68–70].

We note that there are dynamical systems with graphs of such functions as invariant sets. For example, let $g : \mathbb{R} \to \mathbb{R}$ be differentiable, and let $h : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$h(x, t) = (\lambda t, \lambda^{2-s}(x - g)t). \tag{37}$$

Then the graph of f can easily be seen to be a repeller of for h, where f is the function given by

$$f(t) = \sum_{k=1}^{\infty} \lambda^{(s-2)k} g(\lambda^k t). \tag{38}$$

In the following two subsections we prove lower and upper bound on critical order of the Weierstrass function.

5.1 *Lower bound on critical order*

To check the fractional differentiability at any point, say τ, we use $t' = t - \tau$ and $\widetilde{W}_\lambda(t', \tau) = W_\lambda(t' + \tau) - W_\lambda(\tau)$ so that $\widetilde{W}_\lambda(0, \tau) = 0$. We have

$$\begin{aligned}\widetilde{W}_\lambda(t', \tau) &= \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin \lambda^k (t' + \tau) - \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin \lambda^k \tau \\ &= \sum_{k=1}^{\infty} \lambda^{(s-2)k} (\cos \lambda^k \tau \sin \lambda^k t' + \sin \lambda^k \tau (\cos \lambda^k t' - 1)).\end{aligned} \tag{39}$$

Now we take fractional derivative of this with respect to t'.

$$\frac{\mathrm{d}^q \widetilde{W}_\lambda(t', \tau)}{\mathrm{d}t'^q} = \sum_{k=1}^{\infty} \lambda^{(s-2+q)k} \left(\cos \lambda^k \tau \frac{\mathrm{d}^q \sin \lambda^k t'}{\mathrm{d}(\lambda^k t')^q} + \sin \lambda^k \tau \frac{\mathrm{d}^q (\cos \lambda^k t' - 1)}{\mathrm{d}(\lambda^k t')^q} \right). \tag{40}$$

that the series on the right hand side will converge uniformly for $q < 2 - s$, justifying our action of taking the fractional derivative operator inside the sum.

Also as $t' \to 0$ for any k the fractional derivatives in the summation of equation (40) goes to zero. Therefore it is easy to see from this that

$$\mathbb{D}^q W_\lambda(\tau) = \lim_{t' \to 0} \frac{\mathrm{d}^q \widetilde{W}_\lambda(t', \tau)}{\mathrm{d}t'^q} = 0 \quad \text{for} \quad q < 2 - s. \tag{41}$$

This shows that qth local derivative of the Weierstrass function exists and is continuous, at any point, for $q < 2 - s$.

5.2 *Upper bound on critical order*

For $q > 2 - s$, right hand side of the equation (40) seems to diverge. We now prove that the LFD of order $q > 2 - s$ in fact does not exist. To do this we write the Weierstrass function as follows.

$$W_\lambda(t) = \sum_{k=1}^{N} \lambda^{(s-2)k} \sin \lambda^k t + \lambda^{(s-2)N} W_\lambda(\lambda^N t). \tag{42}$$

We now write

$$\frac{\mathrm{d}^q (W_\lambda(t) - W_\lambda(t'))}{\mathrm{d}(t - t')^q} = \sum_{k=1}^{N} \lambda^{(s-2+q)k} \frac{\mathrm{d}^q (\sin(\lambda^k t) - \sin(\lambda^k t'))}{\mathrm{d}(\lambda^k t)^q} + \lambda^{(s-2+q)N} \frac{\mathrm{d}^q (W_\lambda(\lambda^N t) - W_\lambda(\lambda^N t'))}{\mathrm{d}(\lambda^N (t - t'))^q}. \tag{43}$$

We choose N such that $\lambda^{-(N+1)} < |t - t'| \leq \lambda^{-N}$. Now since $|\sin(\lambda^k t) - \sin(\lambda^k t')| \leq \lambda^k |t - t'|$

$$\frac{\mathrm{d}^q |\sin(\lambda^k t) - \sin(\lambda^k t')|}{\mathrm{d}[\lambda^k (t - t')]^q} \leq C \lambda^{k(1-q)} |t - t'|^{1-q}. \tag{44}$$

This implies that the absolute value of the first term in equation (43) is bounded from above by $C\lambda^{(s-2+q)N}$. Now since the first derivative of the Weierstrass function does not exist at any point there exists a sequence of points t_n approaching t' such that $|W_\lambda(\lambda^N t_n) - W_\lambda(\lambda^N t')| \geq c\lambda^N |t_n - t'|$. Therefore there exists a sequence t_n such that

$$\frac{\mathrm{d}^q |W(\lambda^N t_n) - W(\lambda^N t')|}{\mathrm{d}[\lambda^N (t_n - t')]^q} \geq c\lambda^{N(1-q)} |t_n - t'|^{1-q} \tag{45}$$

and this is valid for every c for large enough n. This implies that

$$\lambda^{(s-2+q)N} \frac{\mathrm{d}^q |W_\lambda(\lambda^N t_n) - W_\lambda(\lambda^N t')|}{\mathrm{d}(\lambda^N (t_n - t'))^q} \geq c\lambda^{(s-2+q)N} \tag{46}$$

$$\left|\frac{\mathrm{d}^q(W_\lambda(t_n)-W_\lambda(t'))}{\mathrm{d}(t_n-t')^q}-\lambda^{(s-2+q)N}\frac{\mathrm{d}^q(W_\lambda(\lambda^N t_n)-W_\lambda(\lambda^N t'))}{\mathrm{d}(\lambda^N(t_n-t'))^q}\right|$$
$$\leq \sum_{k=1}^{N}\lambda^{(s-2+q)k}\frac{\mathrm{d}^q|\sin(\lambda^k t_n)-\sin(\lambda^k t')|}{\mathrm{d}(\lambda^k(t_n-t'))^q}. \tag{47}$$

Therefore we get

$$\left|\frac{\mathrm{d}^q(W_\lambda(t_n)-W_\lambda(t'))}{\mathrm{d}(t_n-t')^q}-C\lambda^{(s-2+q)N}\right|\leq C\lambda^{(s-2+q)k}. \tag{48}$$

This implies that

$$\left|\frac{\mathrm{d}^q(W_\lambda(t_n)-W_\lambda(t'))}{\mathrm{d}(t_n-t')^q}\right|\geq C'\lambda^{(s-2+q)k} \tag{49}$$

for $C' > 0$. From this it is clear that the LFD of order greater than $2-s$ does not exist. This concludes the proof.

Summarizing, therefore, the critical order of the Weierstrass function is $2-s$ at all points. It may be noticed that this proof is valid for any $\lambda > 1$. This generalizes a similar result of [4, 36] which is valid only for sufficiently large λ. Thus there is a direct connection between dimension and the differentiability properties for $W_\lambda(t)$. As seen below, this observation is not restricted to $W_\lambda(t)$ but is quite general.

5.3. *Lévy index of a Lévy flights and critical order*

Schlesinger *et al* [71] have considered a Lévy flight on a one dimensional periodic lattice where a particle jumps from one lattice site to other with the probability given by

$$P(x)=\frac{\omega-1}{2\omega}\sum_{j=0}^{\infty}\omega^{-j}[\delta(x,+b^j)+\delta(x,-b^j)], \tag{50}$$

where x is magnitude of the jump, b is a lattice spacing and $b > \omega > 1$. $\delta(x,y)$ is a Kronecker delta. The characteristic function for $P(x)$ is given by

$$\tilde{P}(k)=\frac{\omega-1}{2\omega}\sum_{j=0}^{\infty}\omega^{-j}\cos(b^j k) \tag{51}$$

which is nothing but the Weierstrass function. For this distribution the Lévy index is $\log\omega/\log b$, which can be identified as critical order of $\tilde{P}(k)$.

More generally for the Lévy distribution with index μ the characteristic function is given by

$$\tilde{P}(k)=A\exp c|k|^{\mu}. \tag{52}$$

Critical order of this function at $k=0$ also turns out to be same as μ. Thus the Lévy index can be identified as the critical order of the characteristic function at $k=0$.

We recall the definition of Hölder exponent $h(y)$ of a function f at y [12, 13] as h such that there exists a polynomial $P_n(x)$ of order n, where n is the largest integer smaller than h, that satisfies

$$|f(x) - P_n(x-y)| = O(|x-y|^h), \tag{53}$$

for x in the neighbourhood of y. When h is restricted between 0 and 1 the equation (53) takes the form

$$|f(x) - f(y)| = O(|x-y|^h). \tag{54}$$

The Hölder exponent characterizes the behaviour of the function around a given point. One can study the function at every point and find its pointwise Hölder exponents. According to equation (53) an analytic function has $h = \infty$ at every point.

As can be seen clearly that the definition of the Hölder exponent is not algorithmic and hence methods need to be developed for its determination. In [36] the following result was proved in a slightly different form, which establishes an equivalence between the Hölder exponent and the critical order.

Theorem 5. *The continuous function $f(x)$ has Hölder exponent α, $0 < \alpha < 1$, at y iff $\mathbb{D}^q f(y) = 0$ for all $q < \alpha$ and $\mathbb{D}^q f(y)$ does not exist for $1 > q > \alpha$.*

This result can be understood from the simple example of fractional derivative of a power x^p, given in (5). In this equation if $0 < p < 1$ and $0 < q < 1$ then the RHS exists. But if we take a limit $x \to 0$ then we get

$$\lim_{x\to 0} \frac{d^q x^p}{dx^q} = \begin{cases} 0, & \text{if } q < p \\ \text{constant}, & \text{if } p = q \\ \infty, & \text{otherwise,} \end{cases} \tag{55}$$

The LHS in the above equation is nothing but the LFD of x^p. This shows that the critical order gives the exponent p. This is expected to happen for any function and at any point with a local power law behaviour.

The following corollary follows from the theorem 5 and a well known result giving relation between Hölder exponent and box dimension of a graph of a fractal function [4].

COROLLARY 1

If the critical order of a function $f(x)$ at every point x is α then $\dim_B f = 2 - \alpha$ where $\dim_B f$ is a box dimension of the graph of the function f.

With a slight modification in the proof of theorem 5 a general result giving equivalence between the Hölder exponent and the critical order using the general definition of LFD follows. The functions F and $\widetilde{F}$ are defined in equations (29) and (30) respectively.

Theorem 6. *Let $f : [0,1] \to \mathbb{R}$ be a continuous function.*

(a) *If $\mathbb{D}^q f(y) = 0$ where $N < q < N+1$, for some y, then $h(y) \geq q$.*

(b) *If there exists a sequence* $x_n \to y$ *as* $n \to \infty$ *such that*

$$\lim_{n\to\infty} F(y, x_n - y; q, N) = \pm\infty,$$

for some y*, then* $h(y) \leq q$.

Similarly a following converse of the above theorem can also be proved.

Theorem 7. *Let* $f : [0,1] \to \mathbb{R}$ *be a continuous function.*

(a) *Suppose*

$$|\widetilde{F}_N(x,y)| \leq c|x-y|^\alpha,$$

where $c > 0$, $N < \alpha < N+1$ *and* $|x-y| < \delta$ *for some* $\delta > 0$*. Then* $\mathbb{D}^q f(y) = 0$ *for any* $q < \alpha$ *for* $y \in (0,1)$.

(b) *Suppose that for* $y \in (0,1)$ *and for each* $\delta > 0$ *there exists* x *such that* $|x-y| \leq \delta$ *and*

$$|\widetilde{F}_N(x,y)| \geq c\delta^\alpha,$$

where $c > 0$, $\delta \leq \delta_0$ *for some* $\delta_0 > 0$ *and* $0 < \alpha < 1$*. Then there exists a sequence* $x_n \to y$ *as* $n \to \infty$ *such that*

$$\lim_{n\to\infty} F(y, x_n - y; q, N) = \pm\infty \quad \text{for} \quad q > \alpha.$$

These two theorems give an equivalence between Hölder exponent and the critical order of fractional differentiability. Their proofs are similar to that of theorem 5.

7. Isolated masked singularities

The purpose of this section is to demonstrate the use of LFD to detect masked singularities. We will consider only isolated singularities. We choose the simplest example $f(x) = \sum_{n=0}^{N} a_n x^n + ax^\alpha$, $N < \alpha < N+1$, $x > 0$. Critical order at $x = 0$ gives the order of singularity at that point whereas the value of the LFD $\mathbb{D}^{q=\alpha} f(0)$, viz $a\Gamma(\alpha+1)$, gives strength of the singularity.

Using LFD we can detect a weaker singularity masked by a stronger singularity. As demonstrated below, we can estimate and subtract the contribution due to stronger singularity from the function and find out the critical order of the remaining function. Consider, for example, the function

$$f(x) = \sum_{n=0}^{N} a_n x^n + ax^\alpha + \sum_{n=N+1}^{M} a_n x^n + bx^\beta, \tag{56}$$

where $N < \alpha < N+1 < M < \beta < M+1$ and $x > 0$. LFD of this function at $x = 0$ of the order α is $\mathbb{D}^\alpha f(0) = a\Gamma(\alpha+1)$. Using this estimate of stronger singularity we now write

$$G(x;\alpha) = f(x) - \sum_{n=0}^{N} \frac{f^{(n)}(0)}{\Gamma(n+1)} x^n - \frac{\mathbb{D}^\alpha f(0)}{\Gamma(\alpha+1)} x^\alpha.$$

The critical order of this function, at $x = 0$, is β which is a masked singularity. Notice that the estimation of the weaker singularity was possible in the above calculation just because the LFD gave the coefficient of $x^{\alpha}/\Gamma(\alpha+1)$. This suggests that using LFD, one should be able to extract secondary singularity spectrum masked by the primary singularity spectrum of strong singularities. Hence one can gain more insight into the processes giving rise to irregular behavior.

Comparison of two methods of studying pointwise behavior of functions, one using wavelets and the other using LFD, shows that characterisation of Hölder classes of functions using LFD is direct and involves fewer assumptions. Characterisation of Hölder class of functions with oscillating singularity, e.g. $f(x) = x^{\alpha}\sin(1/x^{\beta})$ $(x > 0, 0 < \alpha < 1$ and $\beta > 0)$, using wavelets needs two exponents [27]. Using LFD, owing to theorems 1 and 2 critical order directly gives the Hölder exponent for such a function.

It has been shown in the context of wavelet transforms that one can detect singularities masked by regular polynomial behavior [12] by choosing an appropriate analysing wavelet. (Wavelets with first n, for some suitable n, moments vanishing are considered appropriate). If one has to extend the wavelet method to unmask weaker singularities, one would then require analysing wavelets with fractional moments vanishing. Notice that one may require this condition along with the condition on first n moments. Further the class of functions to be analysed is in general restricted in these analyses. These restrictions essentially arise from the asymptotic properties of the wavelets used. On the other hand, these restrictions are not relevant while using LFD.

8. Multifractal function

We saw in § 5 that the Weierstrass function is a fractal function, i.e., it has the same Hölder exponent at every point. But there are multifractal functions which have different Hölder exponents at different points. These functions can be used to model various intermitent signals arising in physical applications. Since the critical order gives the local and pointwise behavior of the function, conclusions of the theorems 5, 6 and 7 will carry over even to the case of multifractal functions where we have different Hölder exponents at different points. Selfsimilar multifractal functions have been constructed by Jaffard [13]. We give one specific example of such a function. This function is a solution F of the functional equation

$$F(x) = \sum_{i=1}^{d} \lambda_i F(S_i^{-1}(x)) + g(x), \quad x \text{ real}, \tag{57}$$

where S_i's are the affine transformations of the kind $S_i(x) = \mu_i x + b_i$ (with $|\mu_i| < 1$ and b_i's real) and λ_i's are some real numbers and g is any sufficiently smooth function (it is assumed that g and its derivatives have fast decay). For the sake of illustration we choose $\mu_1 = \mu_2 = 1/3$, $b_1 = 0$, $b_2 = 2/3$, $\lambda_1 = 3^{-\alpha}$, $\lambda_2 = 3^{-\beta}$ $(0 < \alpha < \beta < 1)$ and

$$\begin{aligned} g(x) &= \sin(2\pi x) \quad \text{if} \quad x \in [0,1] \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Such functions are studied in detail in [13] using wavelet transforms where it was shown that the above functional equation (with the parameters we have chosen) has a unique solution F. Further at any point F either has Hölder exponents ranging from α to β or is smooth. A sequence of points $S_{i_1}(0)$, $S_{i_2}S_{i_1}(0), \ldots, S_{i_n}\ldots S_{i_1}(0), \ldots$, where i_k takes values 1 or 2, tends to a point in $[0,1]$ (in fact to a point of a triadic Cantor set) and for the values of μ_is we have chosen this correspondence between sequences and limits is one to one. The solution of the above functional equation is given by [13]

$$F(x) = \sum_{n=0}^{\infty} \sum_{i_1,\ldots,i_n=1}^{2} \lambda_{i_1}\cdots\lambda_{i_n} g(S_{i_n}^{-1}\cdots S_{i_1}^{-1}(x)). \tag{58}$$

Note that with the above choice of parameters the inner sum in (58) reduces to a single term. Jaffard [13] has shown that the local Hölder exponent at y is

$$h(y) = \liminf_{n\to\infty} \frac{\log\left(\lambda_{i_1(y)}\cdots\lambda_{i_n(y)}\right)}{\log\left(\mu_{i_1(y)}\cdots\mu_{i_n(y)}\right)}, \tag{59}$$

where $\{i_1(y), \ldots, i_n(y)\}$ is a sequence of integers appearing in the sum in equation (58) at a point y. It is clear that $h_{\min} = \alpha$ and $h_{\max} = \beta$. The function F at the points of a triadic cantor set have $h(x) \in [\alpha, \beta]$ and at other points it is smooth (where F is as smooth as g). Benzi *et al* [72] have constructed multifractal functions which are random in nature unlike the above nonrandom functions. For still another approach also see [73].

Several well-known 'pathological' functions have been reanalyzed in [29–32] and found to have multifractal nature. Here we consider one example of classical multifractal function

$$R(x) = \sum_{n=1}^{\infty} \frac{1}{n^2}\sin(\pi n^2 x). \tag{60}$$

This function was proposed by Riemann. It turns out that the regularity of this function varies strongly from point to point. Hardy and Littlewood [74] proved that $R(x)$ is not differentiable at x_0 if x_0 is irrational or if x_0 can not be written as $2p+1/2q+1$ $(p, q \in N)$. In fact they showed that the Hölder exponent at these points $\leq 3/4$. Gerver [75] proved the differentiability of $R(x)$ at points of the form $2p+1/2q+1$ $(p, q \in N)$. At these points the Hölder exponent is $3/2$. This function has also been studied in [76, 77]. Jaffard [29] has recently shown that the dimension spectrum of the Riemann function is given as below.

$$\mathrm{d}(\alpha) = \begin{cases} 4\alpha - 2 & \text{if} \quad \alpha \in [\frac{1}{2}, \frac{3}{4}] \\ 0 & \text{if} \quad \alpha = \frac{3}{2} \\ -\infty & \text{otherwise}, \end{cases} \tag{61}$$

where $\mathrm{d}(\alpha)$ gives the Hausdorff dimension of the set when the Hölder exponent is α.

LFD forms one method of studying pointwise behaviour of such multifractal functions alongwith other methods and may considerably reduce the analysis involved. This fact was demonstrated in [36] on a specific example of self-similar multifractal function given by equation (58).

9. Concluding remarks

First we reviewed the applications of various fractional differential equations in different physical situations. It was noted that most of these applications dealt with the asymptotic scaling. Further we reviewed the classical results within the framework of Weyl fractional calculus. However these results were found to be inadequate for the study of pointwise behavior of fractal and multifractal functions. The notion of LFD as developed in [36] was found suitable for this purpose. In terms of the LFD it was also possible to write a fractional Taylor series of a function (useful in an analytic treatment of approximations). It was pointed out that generalization of the notion of tangents to the graph of a function (useful for geometric purposes) is also possible using LFD.

It was established that the critical order of the Weierstrass function is related to the box dimension of its graph. If the dimension of the graph of such a function is $1+\gamma$, the critical order is $1-\gamma$. When γ approaches unity the function becomes increasingly irregular and local fractional differentiability is lost accordingly. Thus there is a direct quantitative connection between the dimension of the graph and the fractional differentiability property of the function. This is one of the remarkable conclusions of the new approach. An important consequence of this approach is that a classification of continuous paths (e.g. fractional Brownian paths) or functions according to local fractional differentiability properties is also a classification according to dimensions (or Hölder exponents).

Also the Lévy index of a Lévy flight on one dimensional lattice is identified as the critical order of the characteristic function of the walk. More generally, the Lévy index of a Lévy distribution is identified as the critical order of its characteristic function at the origin.

We have argued and demonstrated that LFDs are useful for studying isolated singularities and singularities masked by the stronger singularity (not just by regular behavior). It was also shown that the pointwise behavior of irregular (fractal or multifractal) functions can be studied using the methods of this paper.

We note, however, that the treatment of random irregular functions as well as multivariable irregular functions is badly needed. We hope that these problems can be tackled in the near future.

Acknowledgements

One of the authors (KMK) is grateful to CSIR, India for financial assistance and (ADG) is grateful to UGC, India for financial assistance during initial stages of the work. We thank the referee and K R Srinivasan for pointing out some missing references. In particular the referee pointed our attention to the mathematical work [78, 79] dealing with multifractal measures and also informed about the recent work by M Zähle [80, 81].

References

[1] T W Körner, *Fourier Analysis* (Cambridge University Press, Cambridge, 1989)
[2] For historical remarks and construction see, for example [1, 3, 4].
[3] B B Mandelbrot, *The fractal geometry of nature* (Freeman, New York, 1977)

[4] K Falconer, *Fractal geometry* (John Wiley, New York, 1990)
[5] R P Feynmann and A R Hibbs, *Quantum mechanics and path integrals* (McGraw-Hill, New York, 1965)
[6] L F Abott and M B Wise, *Am. J. Phys.* **49**, 37 (1981)
[7] P Constantin, I Procaccia and K R Sreenivasan, *Phys. Rev. Lett.* **67**, 1739 (1991)
[8] P Constantin and I Procaccia, *Nonlinearity* **7**, 1045 (1994)
[9] K Sarkar and C Meneveau, *Phys. Rev.* **E47**, 957 (1993)
[10] J L Kaplan, J Malet-Peret and J A Yorke, *Ergodic Th. Dyn. Syst.* **4**, 261 (1984)
[11] J D Farmer, E Ott and J A Yorke, *Physica* **D7**, 153 (1983)
[12] J F Muzy, E Bacry and A Arneodo, *Phys. Rev.* **E47**, 875 (1993)
[13] S Jaffard, *SIAM J. Math. Anal.* (to appear)
[14] U Frisch and G Parisi, in *Turbulence and predictability in geophysical fluid dynamics and climate dynamics* edited by M Ghil, R Benzi and G Parisi (North-Holland, Amsterdam, 1985)
[15] R Benzi, G Paladin, G Parisi and A Vulpiani, *J. Phys.* **A17**, 3521 (1984)
[16] T C Halsey, M H Jensen, L P Kadanoff, I Procaccia and B I Shraiman, *Phys. Rev.* **A33**, 1141 (1986)
[17] P Collet, J Lobowitz and A Porzio, *J. Stat. Phys.* **47**, 609 (1987)
[18] M H Jensen, L P Kadanoff and I Procaccia, *Phys. Rev.* **A46**, 1409 (1987)
[19] B B Mandelbrot, *Pure Appl. Geophys.* **131**, 5 (1989)
[20] C Meneveau and K R Sreenivasan, *Phys. Rev. Lett.* **59**, 1424 (1987)
[21] A B Chhabra and K R Srinivasan, *Phys. Rev. Lett.* **68**, 2762 (1992)
[22] J Feder, *Fractals* (Pergamon, New York, 1988)
[23] T Vicsek, *Fractal growth phenomenon* (World Scientific, Singapore, 1989)
[24] R Cawley and R D Mauldin, *Adv. Math.* **92**, 192 (1992)
[25] M Holschneider, *J. Stat. Phys.* **77**, 807 (1994)
[26] S Jaffard, in *Wavelets and Applications* edited by Y Meyer (Springer-Verlag, Berlin, 1992)
[27] A Arneodo, E Bacry and J F Muzy, *Phys. Rev. Lett.* **74**, 4823 (1995)
[28] S Jaffard and Y Meyer, *Memoirs of the AMS* (to appear)
[29] S Jaffard, *Preprint*
[30] S Jaffard and B B Mandelbrot, *Adv. Math.* (to appear)
[31] I Daubechies and J Lagarias, *Rev. Math. Phys.* **6**, 1033 (1994)
[32] S Jaffard, *Preprint*
[33] K B Oldham and J Spanier, *The fractional calculus* (Academic Press, New York, 1974)
[34] K S Miller and B Ross, *An Introduction to the fractional calculus and fractional differential equations* (John Wiley, New York, 1993)
[35] B Ross, in *Fractional calculus and its applications: Lecture notes in mathematics* (Springer, New York, 1975) vol. 457, p. 1
[36] K M Kolwankar and A D Gangal, *Chaos* **6**, 505 (1996)
[37] R Hilfer, *Phys. Scr.* **44**, 321 (1991)
[38] R Hilfer, *Phys. Rev. Lett.* **68**, 190 (1992)
[39] K M Kolwankar, *Preprint*
[40] T F Nonnenmacher, *J. Phys.* **A23**, L697 (1990)
[41] M Giona and H E Roman, *J. Phys.* **A25**, 2093 (1992)
[42] H E Roman and M Giona, *J. Phys.* **A25**, 2107 (1992)
[43] N Patzschke and M Zähle, *Stochastic Process Appl.* **43**, 165 (1992)
[44] B B Mandelbrot and J W Van Ness, *SIAM Rev.* **10**, 422 (1968)
[45] W G Glöckle and T F Nonnenmacher, *J. Stat. Phys.* **71**, 741 (1993)
[46] M F Schlesinger, *J. Stat. Phys.* **36**, 639 (1984)
[47] K L Sebastian, *J. Phys.* **A28**, 4305 (1995)
[48] J P Bouchaud and A Georges, *Phys. Rep.* **195**, 127 (1990)
[49] M F Shlesinger, B J West and J Klafter, *Phys. Rev. Lett.* **58**, 1100 (1987)
[50] M F Shlesinger, G M Zaslavsky and J Klafter, *Nature (London)* **363**, 31 (1993)
[51] M F Shlesinger *et al* (eds), *Lévy flights and related topics in physics* (Springer, Berlin, 1995)
[52] T H Solomon, E R Weeks and H L Swinney, *Phys. Rev. Lett.* **71**, 3975 (1993)
[53] A Ott, J P Bouchaud, D Langvin and W Urbach, *Phys. Rev. Lett.* **65**, 2201 (1990)

[54] W Wyss, *J. Math. Phys.* **27**, 2782 (1986)
[55] W R Schneider and W Wyss, *J. Math. Phys.* **30**, 134 (1989)
[56] G Jumarie, *J. Math. Phys.* **33**, 3536 (1992)
[57] H C Fogedby, *Phys. Rev. Lett.* **73**, 2517 (1994)
[58] G M Zaslavsky, *Physica* **D76**, 110 (1994)
[59] A Zygmund, *Trignometric series* 2nd ed. (Cambridge Univ. Press, New York, 1959) vols I, II
[60] E M Stein, *Singular integrals and differentiability properties of functions* (Princeton University press, Princeton, 1970)
[61] E M Stein and A Zygmund, *Proc. London Math. Soc.* **A14**, 249 (1965)
[62] G V Welland, *Proc. Am. Math. Soc.* **19**, 135 (1968)
[63] I Eyink, *J. Stat. Phys.* **78**, 353 (1995)
[64] S Jaffard, *SIAM J. Math. Anal.* (to appear)
[65] T J Osler, *SIAM J. Math. Anal.* **2**, 37 (1971)
[66] R D Mauldin and S C Williams, *Trans. Am. Math. Soc.* **298**, 793 (1986)
[67] B R Hunt, Preprint (1996)
[68] G H Hardy, *Trans. Am. Math. Soc.* **17**, 301 (1916)
[69] A S Besicovitch and H D Ursell, *J. Lond. Math. Soc.* **12**, 18 (1937)
[70] M V Berry and Z V Luwis, *Proc. R. Soc.* **A370**, 459 (1980)
[71] M F Shlesinger, *Physica* **D38**, 304 (1989)
[72] R Benzi, L Biferale, A Crisanti, G Paladin, M Vergassola and A Vulpiani, *Physica* **D65**, 352 (1993)
[73] A Juneja, D P Lathrop, K R Sreenivasan and G Stolovitzky, *Phys. Rev.* **E41**, 5179 (1994)
[74] G H Hardy and E Littlewood, *Acta Math.* **37**, 194 (1914)
[75] J Gerver, *Am. J. Math.* **93**, 33 (1970)
[76] M Holschneider and Ph Tchamitchian, *Invent. Math.* **105**, 157 (1991)
[77] J J Duistermaat, *Overdruk* **9**, 303 (1991)
[78] L Olsen, *Pitman research notes in mathematica series* **307**, (1994)
[79] R Riedi, *J. Math. Anal. Appl.* **189**, 462 (1995)
[80] M Zähle, Friedrich-Schiller Universität Jena, Preprint (1996)
[81] M Zähle, *Fractals* **3**, 747 (1995)

Signature of chaos in power spectrum

M C VALSAKUMAR, S V M SATYANARAYANA and V SRIDHAR
Theoretical Studies Section, Materials Science Division, Indira Gandhi Centre for Atomic Research, Kalpakkam 603 102, India

Abstract. We investigate the nature of the numerically computed power spectral density $P(f, N, \tau)$ of a discrete (sampling time interval, τ) and finite (length, N) scalar time series extracted from a continuous time chaotic dynamical system. We highlight how $P(f, N, \tau)$ differs from the rue power spectrum and from the power spectrum of a general stochastic process. Non-zero τ leads o aliasing; $P(f, N, \tau)$ decays at high frequencies as $[\pi\tau / \sin \pi\tau f]^2$, which is an aliased form of the $1/f^2$ decay. This power law tail seems to be a characteristic feature of all continuous time dynamical systems, chaotic or otherwise. Also the tail vanishes in the limit of $N \to \infty$, implying hat the true power spectral density must be band width limited. In striking contrast the power spectrum of a stochastic process is dominated by a term independent of the length of the time series at all frequencies.

Keywords. Chaos; time series analysis; Power spectrum; Aliasing.

PACS Nos 05.45; 47.52

1. Introduction

The problem of distinguishing chaos from noise as the mechanism responsible for the irregular oscillations observed in a scalar time series is of great importance owing to its relevance to a variety of experimental contexts [1]. If the erratic oscillations are chaotic, then there exists a low dimensional deterministic dynamics responsible for it. Then, in principle, it is possibile to reconstruct the relevant dynamical equations, paving way for a complete understanding of the phenomenon. There has been an intense activity in this direction resulting in the formulation of several numerical approaches based on metric and topological considerations, see for example [2]. In this paper we shall focus attention on the analysis of the spectral measure – one of the traditional tools employed in the study of stationary stochastic processes as well as deterministic dynamical systems. For example, a periodic system with frequency f has a pure point spectral measure with a spectral density (henceforth called power spectrum) constituted by sharp(delta) peaks at f, and all its harmonics, in general. On the other hand, the spectral measure of a *generic* stationary stochastic process has an *absolutely continuous* part and a corresponding broad power spectrum. Chaotic time series also leads to a broadened power spectrum and hence it becomes difficult to distinguish chaos from noise. Intuitively, it is clear that finite changes can not take place in a deterministic system linear or nonlinear (chaotic or otherwise), in an inifinitesimal interval of time, whereas finite changes are possible in any interval of time in a generic stochastic process. This argument holds good even for cases of deterministic dynamical systems with non differentiable solutions, like the kicked

$C(f)$ defined in (2) and (3), henceforth, will be referred to as the true power spectrum in this paper.

In general, the computer or experimental realization of state variable $x(t)$ is in the form of a scalar discrete time series. This is equivalent to obtaining a time series by sampling $x(t)$ with a non zero sampling time τ, up to a finite length N, giving $\{x_j = x(t = j\tau); j = 0, N-1\}$. We now proceed to define the power spectrum $P(f, N, \tau)$ of such a discrete scalar time series such that it agrees with $C(f)$ in the limit $\tau \to 0$ and $N \to \infty$. The discrete version of the autocorrelation is defined as

$$c_j(N) = \left\langle \frac{1}{N} \sum_{l=0}^{N-1-|j|} x_l x_{l+|j|} \right\rangle, \tag{4}$$

where $\langle\cdot\rangle$ represents the average over several initial conditions and trajectories. This averaging is performed so as to ensure that the autocorrelation function of the discrete time series $c_j(N)$ is identical to that of the continuous time process $c(t)$ evaluated at $t = j\tau$ in the limit $N \to \infty$. $P(f, N, \tau)$ is then defined as

$$P(f, N, \tau) = \tau \sum_{j=-(N-1)}^{(N-1)} c_j(N) \mathrm{e}^{i2\pi f\tau j}. \tag{5}$$

$P(f, N, \tau)$ can also be represented in the form analogous to equation (3). Let $X(f, N, \tau)$ be the discrete Fourier transform of $\{x_j\}$,

$$X(f, N, \tau) = \sum_{j=0}^{N-1} x_j \, \mathrm{e}^{i2\pi f\tau j} \tag{6}$$

and the corresponding power spectrum is

$$P(f, N, \tau) = \frac{\tau}{N} \langle |X(f, N, \tau)|^2 \rangle. \tag{7}$$

In this paper we refer to $P(f, N, \tau)$ as the computed power spectrum or simply the power spectrum. The computed power spectrum $P(f, N, \tau)$ is equal to the true power spectrum $C(f)$ in the following limit.

$$C(f) = \lim_{\tau \to 0} \lim_{N \to \infty} P(f, N, \tau). \tag{8}$$

In our numerical investigations, we used the definitions given in equations (6) and (7) to compute the power spectrum of chaotic time series. Power spectra of different segments of length N, extracted from a time series of the dynamical system corresponding to a given initial condition, are computed. The computation is repeated for different initial conditions. An average over this ensemble of power spectra is carried out to obtain $P(f, N, \tau)$.

The power spectra of chaotic dynamical systems are found to exhibit an exponential decay followed by a much slower decay (like an algebraic decay). The power spectra of two representative models exhibiting chaos, namely, the Lorenz and Rossler models, are shown in figure 1(a) and figure 1(b), respectively. The nonexponential tail, though appears like an algebraic decay, does not fit to any power law. But, surprisingly, we find

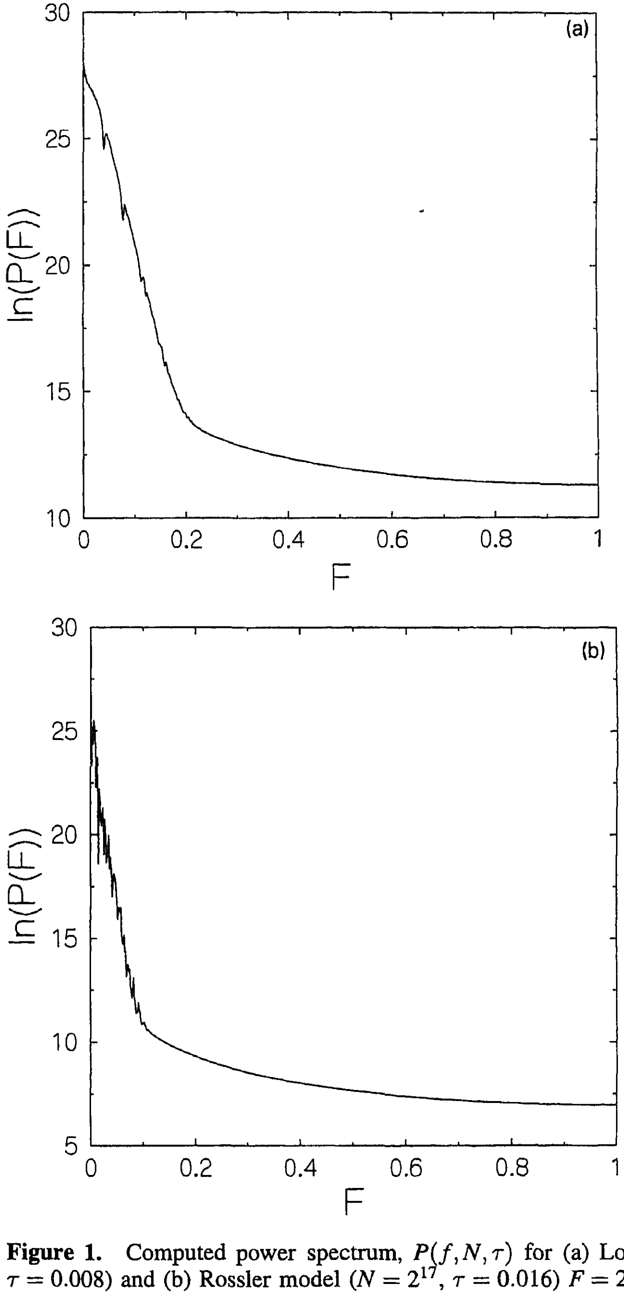


Figure 1. Computed power spectrum, $P(f, N, \tau)$ for (a) Lorenz model ($N = 2^{17}$, $\tau = 0.008$) and (b) Rossler model ($N = 2^{17}$, $\tau = 0.016$) $F = 2\tau f$.

that the tail exhibits a universal $[\pi\tau / \sin \pi\tau f]^2$ decay. Figures 2(a) – 2(f) show the power spectral density in the nonexponential region to be proportional to $[\pi\tau / \sin \pi\tau f]^2$. It is obvious from the data that, without exception, the fit is excellent. The origin of such a functional form for the computed power spectrum at high frequencies will form the

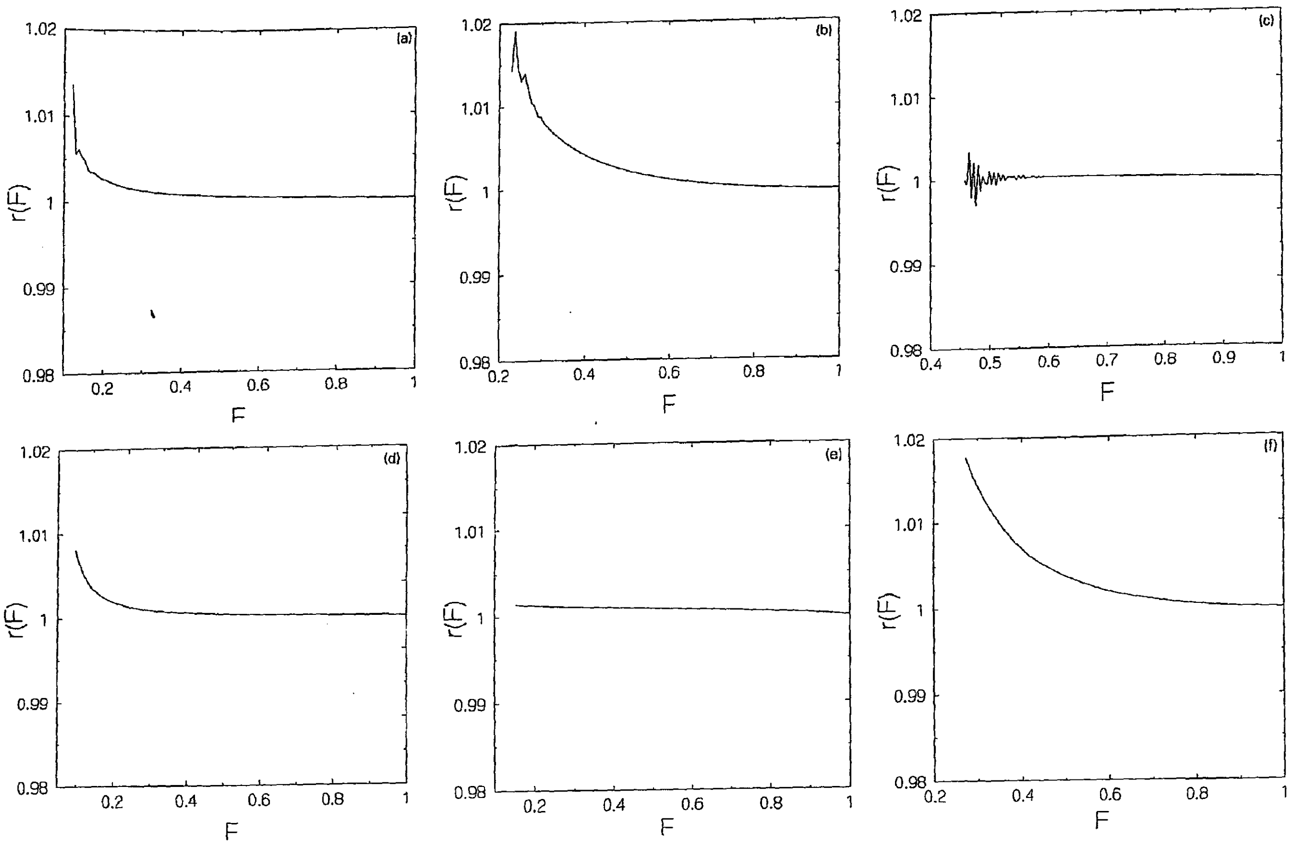


Figure 2. The ratio of the computed power spectrum and the aliased $1/f^2$ spectrum at high frequency region, $r(F) = P(f,N,\tau)/\mathrm{cosec}^2((\pi/2)F)$ where $F = f/f_m = 2\tau f$ for **(a)** Rossler model, **(b)** Lorenz model (quasi-periodic route), **(c)** Raman model, **(d)** Lorenz model (intermittency route), **(e)** Anantha oscillator and **(f)** Quasiperiodically kicked oscillator. These results are found to be independent of the choice of τ and N.

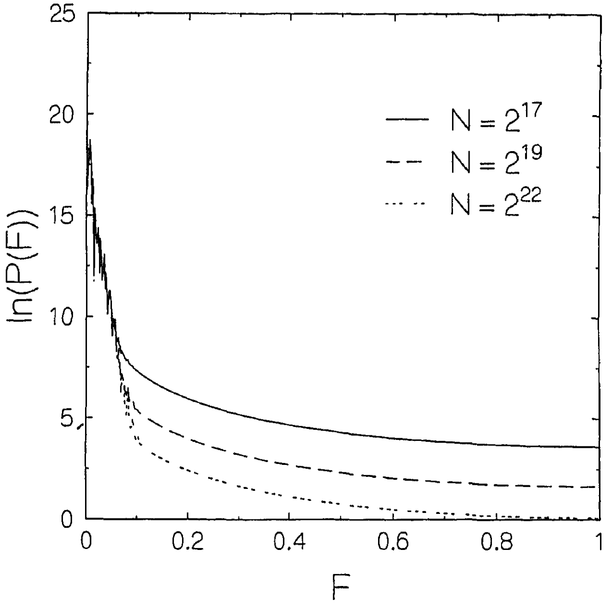


Figure 3. The scaling of the computed power spectrum with the size of the discrete time series, for Rossler oscillator.

subject matter of the next two sections. We observe that the spectral density in the nonexponential tail region is smooth and that it scales as $1/N$. The N-dependence of the power spectral density at a given frequency in the tail region is shown in figure 3. Detailed description of N-dependence is presented in § 5.

We have carried out the calculations in double and quadruple precision to see the influence of machine accuracy on our results. We find that our conclusions remain the same.

3. Relation between the true and computed power spectra

In order to understand the findings reported in the last section in a proper perspective, we derive here a formal relation between the computed and true power spectra. In doing this, we first define the following two functions, namely

$$w_N(t) = \begin{cases} 1 & \text{for } \quad -(N-1)\tau \le t \le (N-1)\tau \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

and

$$\hat{c}(t) = \tau \sum_{j=-\infty}^{\infty} c(t)\delta(t - j\tau). \tag{10}$$

The correlation function, in general, depends on N and it can be rewritten as

$$c_j(N) = c_j + d_j(N), \tag{11}$$

where $c_j = c(t = j\tau)$ and $d_j(N)$ is a remainder which vanishes in the limit $N \to \infty$. This implies that $P(f,N,\tau)$ can be expressed as a sum of two terms

$$P(f,N,\tau) = P_1(f,N,\tau) + P_2(f,N,\tau), \tag{12}$$

where

$$P_1(f,N,\tau) = \int_{-\infty}^{\infty} \mathrm{d}t\, \hat{c}(t) w_N(t) \exp(i2\pi ft), \tag{13}$$

and $P_2(f,N,\tau)$, related to $\{d_j(N)\}$, is the remainder that vanishes in the limit $N \to \infty$. $P_1(f,N,\tau)$ given in (13) can be rewritten in terms of the functions defined in equations (9) and (10) as

$$P_1(f,N,\tau) = \int_{-\infty}^{\infty} \mathrm{d}t\, \hat{c}(t) w_N(t) \mathrm{e}^{i2\pi ft}. \tag{14}$$

Using the convolution theorem, we can write

$$P_1(f,N,\tau) = \int_{-\infty}^{\infty} \mathrm{d}f' \hat{C}(f-f') W_N(f'), \tag{15}$$

where $\hat{C}(f)$ and $W_N(f)$ are the Fourier transforms of $\hat{c}(t)$ and $w_N(t)$, respectively. The Fourier transform of $w_N(t)$ is given by

$$W_N(f) = \frac{\sin 2\pi(N-1)f\tau}{\pi f}. \tag{16}$$

This can be recast as

$$W_N(f) = \delta(f) + Y_N(f), \tag{17}$$

where $\lim_{N\to\infty} Y_N(f) = 0$. Substituting (17) in (15) gives

$$P_1(f,N,\tau) = \hat{C}(f) + Q_N(f,N,\tau) \tag{18}$$

such that $\lim_{N\to\infty} Q_N(f,N,\tau) = 0$. Using (12) and (18), we can write the computed power spectrum as

$$P(f,N,\tau) = \hat{C}(f) + R(f,N,\tau), \tag{19}$$

where $R(f,N,\tau)$ vanishes in the limit $N \to \infty$. Using once again convolution theorem, the Fourier transform of $\hat{c}(t)$ can be written as

$$\hat{C}(f) = \int_{-\infty}^{\infty} \mathrm{d}f' C(f-f') \sum_{m=-\infty}^{\infty} \delta\left(f' + \frac{m}{\tau}\right). \tag{20}$$

This can be simplified and we get

$$\hat{C}(f) = \sum_{m=-\infty}^{\infty} C\left(f + \frac{m}{\tau}\right). \tag{21}$$

Finally using (19) and (21), we get the desired result.

$$P(f,N,\tau) = \sum_{m=-\infty}^{\infty} C\left(f + \frac{m}{\tau}\right) + R(f,N,\tau), \tag{22}$$

where $P(f,N,\tau)$ is the computed power spectrum and $C(f)$ is the true power spectrum. This equation formally relates the computed power spectrum to the true one. It brings out clearly the role played by τ and N. The first term represents $\lim_{N\to\infty} P(f,N,\tau)$ and the second term reflects the finite size effects. We shall be analysing in the following sections, how a relation of this kind explains the results.

4. Aliasing

Power spectrum of any sampled data, with an interval of sampling τ being nonzero, is defined only in the domain $-f_m \leq f \leq f_m$, where $f_m = 1/2\tau$, is the Nyquist frequency. In sampling a given function $x(t)$, τ can be tuned. $\tau \to 0$ corresponds to sampling the entire function. But, for a fixed τ, the signal might contain frequencies higher than the Nyquist frequency. In such a case, the spectral densities corresponding to the frequencies larger than f_m will be folded into the region $|f| < f_m$. This is the phenomenon of aliasing, see for example [13].

The first term of (22) which represents $\lim_{N\to\infty} P(f,N,\tau)$ is an aliased version of true power spectrum $C(f)$. This clearly establishes the role played by nonzero τ resulting in the change in the functional form of the decay in the power spectrum due to aliasing. The remainder $R(f,N,\tau)$ in (22) depends both on τ and N. This term, owing to its dependence on τ, could itself, in principle, be an aliased version of some other function. We derive, in this section, the aliased versions of power spectra with (i) an exponential and (ii) a power law dependence on f, and discuss the implications with reference to the observed numerical results.

4.1 *Aliased exponential decay*

We consider a power spectrum with an exponential dependence on f.

$$C(f) = a\mathrm{e}^{-bf}. \tag{23}$$

The aliased version $P_a(f)$ of $C(f)$ can be written as

$$P_a(f) = \sum_{j=-\infty}^{\infty} C(f + 2jf_m), \tag{24}$$

where $f_m = 1/2\tau$, the Nyquist frequency. The sum can be easily evaluated and we get

$$P_a(F) = a\frac{\cosh[bf_m(1-F)]}{\sinh[bf_m]}, \tag{25}$$

where $F = f/f_m$, is the scaled frequency. It is worth mentioning that, as a function of the scaled frequency F, the discrepancy between the aliased and the true power spectra gets confined to a region closer to the end point as τ decreases.

4.2 *Aliasing of power law decay*

Here we first derive an expression for the aliased version of $1/f^2$ decay and then general expression for any even power of $1/f$. The true power spectrum is given by

$$C(f) = \frac{A}{f^2}. \tag{26}$$

The aliased form can be written as

$$P_a(F) = \frac{A}{f_m^2} \sum_{j=-\infty}^{\infty} \frac{1}{(2j+F)^2}. \tag{27}$$

Using the integral representation

$$\frac{1}{F^2} = \int_0^\infty \mathrm{d}t\, t\, \mathrm{e}^{-Ft} \tag{28}$$

we can rewrite $P_a(f)$ as

$$\begin{aligned} P_a(F) &= \frac{A}{f_m^2} \int_0^\infty \mathrm{d}t\, t\, \frac{\cosh(1-F)t}{\sinh t} \\ &= -\frac{A}{2f_m^2} \frac{\partial}{\partial F} \int_{-\infty}^\infty \mathrm{d}t\, \frac{\sinh(1-F)t}{\sinh t} \qquad (29) \\ &= -\frac{A}{2f_m^2} \frac{\partial}{\partial F} J(F). \end{aligned}$$

Using contour integration J can be evaluated as

$$J = \pi \cot \frac{\pi F}{2} \tag{30}$$

and

$$P_a(F) = \frac{A\pi^2}{4f_m^2} \operatorname{cosec}^2 \frac{\pi F}{2}. \tag{31}$$

The above can be recast in the following convenient form

$$P_a(f) = A \left[\frac{\pi\tau}{\sin \pi\tau f}\right]^2. \tag{32}$$

From (32), it is obvious that $\lim_{\tau\to 0} P_a(f) = C(f)$. However, for any finite τ, $P_a(F)$ is shape invariant and the discrepancy between the computed and the true power spectra does not vanish at any frequency. This functional form fits extremely well the power spectra of chaotic systems in the high frequency (nonexponential) region as can be seen from the figures 2(a)–2(f). This implies that the power spectra of chaotic time series have a universal $1/f^2$ tail, which is seen as a $[\pi\tau/\sin \pi\tau f]^2$ tail due to aliasing.

The aliased version of power law decay with an even exponent (say, $2l$) can also be exactly calculated.

$$P_a(F) = \frac{A}{f_m^{2l}} \sum_{j=-\infty}^{\infty} \frac{1}{(2j+F)^{2l}}. \tag{33}$$

This can be further rewritten as

$$P_a(F) = -\frac{A}{2f_m^{2l}} \frac{1}{(2l-1)!} \frac{\partial^{2l-1}}{\partial F^{2l-1}} \int_{-\infty}^{\infty} \mathrm{d}t \, \frac{\sinh(1-F)t}{\sinh t}. \tag{34}$$

Since the integral in (34) is already evaluated (see (30)), $P_a(F)$ can easily be obtained by computing the appropriate derivative. in (34).

5. Finite size effects

As has been emphasized at the beginning, the finite size of the time series also leads to a discrepancy between the computed and the true power spectra. The formal relation between the computed and the true power spectra described in (22) of § 3, has a part that depends on N, denoted by $R(f, N, \tau)$. This remainder is such that it vanishes in the limit $N \longrightarrow \infty$.

Numerical investigation of the power spectral density of several chaotic dynamical systems listed in table 1, show that there is an exponential decay up to some frequency $f_\star$ followed by a $[\pi\tau/\sin \pi\tau f]^2$ tail. Our studies on the scaling of the power spectral density of the chaotic dynamical systems with respect to the system size N reveals that the spectral density corresponding to the frequencies in the exponential decay region is independent of N, whereas the ones corresponding to the frequencies in the tail region scale as $1/N$. The scaling of the power spectrum with N for Rossler is presented in figure 3. Since, by definition, see (22), only $R(f, N, \tau)$ can depend on N, it is clear from (32), together with the numerical results, that

$$R(f, N, \tau) \sim \frac{1}{N} \left[\frac{\pi\tau}{\sin \pi\tau f}\right]^2 \tag{35}$$

to the leading order in $1/N$. Since $R(f, N, \tau)$ accounts for the whole of the tail (see the fit in figures 2(a)–2(f)), within the limits of the present numerical analysis, it is evident that $\hat{C}(f)$ does not have any spectral component in the tail region. This implies that the power spectrum is band width limited.

In the experimental context, the interest is mainly to distinguish chaotic oscillations from random oscillations (or stochastic signals). The band width limited character of the power spectrum signals the presence of chaotic oscillations in the time series. In a generic noise, we do not expect the power spectrum of such a process to be band width limited. For example the power spectrum of white noise is broad and is independent of N. One can understand that in a determinisitic dynamical system, chaotic or otherwise, no drastic changes can take place in an extremely short interval of time and hence the power spectral density in the high frequency region will accordingly be zero. This results in a band width limited power spectrum. On the other hand, by definition, a generic stochastic process is one in which changes occur on any time scale. This leads to the conclusion that the power spectral density of such a process must be nonzero for all the frequencies. In Appendix II, we present a proof that for a stationary stochastic process with absolutely continuous spectral measure, the finite time power spectrum is dominated by an N independent term at all frequencies.

The $1/N$ scaling, together with an aliased $1/f^2$ decay at high frequency preceded by an exponential decay in the power spectrum single out chaotic time series from periodic, quasiperiodic and noisy signals.

6. Critical remarks on existing theoretical arguments

In [3–5] the authors present a case for the exponential decay of the power spectra of chaotic systems. The argument rests on the following definition of the power spectrum

$$C(f) = |X(f)|^2 \tag{36}$$

with

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{i2\pi ft}\mathrm{d}t. \tag{37}$$

By extending $x(t)$ to the complex time domain, the authors argue that $X(f)$ can be represented as a sum of exponentials, and hence $C(f)$ must be a sum of exponentials. They have tacitly assumed the existence of $X(f)$.

In this section we present the problems associated with the above argument. The proper definition of $C(f)$ is, see [6]

$$C(f) = \lim_{T\to\infty} \frac{1}{T} |X_T(f)|^2 \tag{38}$$

with

$$X_T(f) = \int_0^T x(t)e^{2\pi ft}\mathrm{d}t. \tag{39}$$

From the numerical simulations we know that the spectral measure of chaotic systems is predominantly absolute-continuous at high frequencies. We find that the associated power spectral density $C(f)$ is nonzero and bounded in the high frequency domain. But, the presence of the factor $1/T$ in (38) necessitates vanishing of $C(f)$ whenever $X(f)$ takes on a nonzero, but finite value. Since $C(f)$ of the system under investigation is nonzero, it is clear that $X(f) = \lim_{T\to\infty} X_T(f)$ can not exist. The finite time Fourier transform $X_T(f)$ should diverge as $\sqrt{T}$ as $T \to \infty$ (our $X(f, N, \tau)$ diverges as $\sqrt{N}$). Clearly these authors have neglected this part of the integral and concentrated only on the principal value, which in fact does not contribute to $C(f)$ at all.

In Appendix I, we show that the power spectra of the deterministic dynamical systems whose solution is smooth and bounded (which serves as sufficiency condition but not necessary) will have a $1/f^2N$ tail.

The computed power spectrum $P(f,N,\tau)$ differs from the true one $S(f)$ due to nonzero sampling time τ and finite length of the time series N. The nonzero sampling time τ gives rise to aliasing, which changes the functional form of the decay. The computed power spectrum decays as $(\cosh bf_m(1-F))/\sinh bf_m$ up to a frequency f_c and subsequently decays as $1/N[\pi\tau/\sin\pi\tau f]^2$ for $f > f_c$. These expressions are calculated to be the aliased versions of e^{-bf} and $1/Nf^2$ respectively, which means the true power spectrum is band width limited. The feature of $1/Nf^2$ decay is observed in all the chaotic dynamical system studied. The finite size N of the time series results in contributing a spurious tail to an otherwise band width limted power spectrum. This feature turns out to be useful for distinguishing chaos from a generic stationary stochastic process as the former scales as $1/N$ while the latter is independent of N.

Appendix I

We now proceed to prove the statement that the power spectrum, defined for the signal in the interval $[0,T]$, of a bounded function belonging to C^∞ class shows a $(f^2T)^{-1}$ tail for $|f| > f_c$.

Consider a dynamical system

$$\frac{dx_i(t)}{dt}=f_i(\mathrm{x}(t)),\quad i=1,\ldots,d \tag{A1}$$

with the initial conditions

$$x_i(0)=x_{i0}. \tag{A2}$$

The formal solution of the above equations is given by

$$x_j(t)=x_j(\mathrm{x}_0,t)=e^{tL}x_{j0}, \tag{A3}$$

where

$$L=\sum_{i=1}^{d} f_i(\mathrm{x}_0)\,\frac{\partial}{\partial x_{i0}}. \tag{A4}$$

The finite time Fourier transform of $x_j(t)$ is defined as

$$X_{j,T}(f)=\int_0^T dt x_j(\mathrm{x}_0,t)e^{i2\pi ft}. \tag{A5}$$

Substituting (A3) in (A5) and simplifying, we get

$$X_{j,T}(f)=\frac{1}{i2\pi f}\left[1-\frac{iL}{2\pi f}\right]^{-1}[e^{i2\pi fT}e^{LT}x_{j0}-x_{j0}]. \tag{A6}$$

The power spectrum $P_{j,T}(f)=1/T|X_{j,T}(f)|^2$ is given by

$$P_{j,T}(f)=\frac{1}{4\pi^2f^2T}\left|\left[1-\frac{iL}{2\pi f}\right]^{-1}[e^{i2\pi fT}e^{LT}x_{j0}-x_{j0}]\right|^2=\frac{|Q|^2}{4\pi^2f^2T}. \tag{A7}$$

The expression for Q can be written as

$$Q = \mathrm{e}^{i2\pi f T}\sum_{n=0}^{\infty}\left(\frac{i}{2\pi f}\right)^{n} L^{n}x_j(\mathrm{x}_0, T) - \sum_{n=0}^{\infty}\left(\frac{i}{2\pi f}\right)^{n} L^{n}x_{j0}. \tag{A8}$$

From (A8) we get

$$|Q| \leq \sum_{n=0}^{\infty}\left(\frac{1}{|2\pi f|^{n}}\right)(|L^{n}x_j(\mathrm{x}_0, T)| + |L^{n}x_{j0}|). \tag{A9}$$

From the formal solution (A3) it is easy to see that

$$L^{n}x_j(\mathrm{x}_0, T) = \frac{\mathrm{d}^n}{\mathrm{d}t^n}x_j(\mathrm{x}_0, t)|_{t=T} \quad \text{and} \quad L^{n}x_{j0} = \frac{\mathrm{d}^n}{\mathrm{d}t^n}x_j(\mathrm{x}_0, t)|_{t=0}. \tag{A10}$$

Assuming that the trajectory $x_j(\mathrm{x}_0, t)$ of the dynamical system (A1) belongs to C^{∞} class, it follows that there exists an upper bound M_n for $|\mathrm{d}^n/\mathrm{d}t^n x_j(\mathrm{x}_0, t)|, n \geq 1$. Hence we obtain

$$|L^{n}x_j(\mathrm{x}_0, t)| \leq M_n \quad \text{for all} \quad t \in [0, \infty). \tag{A11}$$

We define a parameter $M = \sup_n (M_n)^{(1/n)}$. It can be seen that

$$\sum_{n=1}^{\infty}\frac{|L^{n}x_j(\mathrm{x}_0, t)|}{|2\pi f|^{n}} \leq \sum_{n=1}^{\infty}\left(\frac{M}{|2\pi f|}\right)^{n} \tag{A12}$$

for all $t \in [0, \infty)$. The above series is convergent for $|2\pi f| > M(= 2\pi f_c)$. The series can then be summed to get

$$|Q| \leq 2\left[M_0 + \frac{M/|2\pi f|}{1 - M/|2\pi f|}\right]. \tag{A13}$$

This illustrates that for $|2\pi f| > M(= 2\pi f_c)$, the power spectrum will have a tail

$$P_{j,T}(f) \sim \frac{M_0^2}{(\pi f)^2 T}. \tag{A14}$$

This will manifest as a $(\pi\tau/\sin\pi f\tau)^2$ tail for a discrete time series sampled at a time step τ (due to aliasing), consistent with our numerical results for the power spectra of chaotic dynamical systems.

Appendix II

Let $x(t)$ be a stationary stochastic process with an absolutely continuous spectral measure. Let $\{x_j = x(t = j\tau);\ j = 0, \ldots, N-1\}$ be a discretely sampled data set. $P(f, N, \tau)$ can be written as

$$P(f, N, \tau) = \frac{\tau}{N}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}\langle x_j x_k\rangle \mathrm{e}^{i2\pi f\tau(j-k)}. \tag{A1}$$

ince the process is stationary one can write $\langle x_j x_k \rangle = c(\tau(j-k))$. Substituting this and implifying, we get

$$P(f,N,\tau) = \int_{-\infty}^{\infty} \mathrm{d}f' C(f') S_N \,, \tag{A2}$$

where

$$S_N = \tau\left[1 + 2\sum_{l=1}^{N-1}\left(1-\frac{l}{N}\right)\cos 2\pi\tau(f-f')l\right] \tag{A3}$$

nd $C(f)$ is the true (unaliased) power spectrum. $P(f,\ N,\ \tau)$ can be equivalently epresented as (see eq. (22))

$$P(f,N,\tau) = \sum_{m=-\infty}^{\infty} C\left(f+\frac{m}{\tau}\right) + R(f,N,\tau). \tag{A4}$$

Thus it is only the true spectral density $C(f)$ that goes as input to the analysis. By definition, a generic stationary stochastic process is characterized by a spectral measure which is absolutely continuous, and is nonzero for all frequencies. It is clear from (A4) that if the true power spectrum is nonzero, so is the aliased power spectrum. This implies that for a given $N, P(f,N,\tau)$ of a generic stationary stochastic process consists of both an N-independent term and terms depending on N. In the following analysis we calculate the N dependent terms explicitly to understand which term dominates in the large N limit.

Let $z = 2\pi(f-f')\tau$, then

$$S_N = \tau\left[1 + 2\mathrm{Re}\,\frac{\mathrm{e}^{iNz}-1}{\mathrm{e}^{iz}-1} - \frac{2}{N}\frac{\partial}{\partial z}\mathrm{Im}\,\frac{\mathrm{e}^{iNz}-1}{\mathrm{e}^{iz}-1}\right]. \tag{A5}$$

This can be evaluated as

$$S_N = \frac{\tau}{N}\frac{\sin^2(Nz/2)}{\sin^2(z/2)}. \tag{A6}$$

The power spectrum now takes the form

$$P(f,N,\tau) = \int_{-\infty}^{\infty} \mathrm{d}f' C(f') \frac{\tau}{N}\frac{\sin^2 N\pi\tau(f-f')}{\sin^2 \pi\tau(f-f')}. \tag{A7}$$

Noting that S_N is periodic in $1/\tau$, the above integral can be written as

$$P(f,N,\tau) = \sum_{l=-\infty}^{\infty} \frac{1}{\pi}\int_{-(\pi/2)}^{\pi/2} \mathrm{d}\xi C\left(f+\frac{l}{\tau}+\frac{\xi}{\pi\tau}\right)\frac{\sin^2 N\xi}{N\sin^2\xi}. \tag{A8}$$

Since the spectral measure is absolutely continuous, the power spectrum $C(f)$ is smooth and one can expand $C(f)$ as a Taylor's series in $\xi/\pi\tau$, to give the power spectrum the following form

$$P(f,N,\tau) = \sum_{l=-\infty}^{\infty}\left[\sum_{m=0}^{\infty} C^{(m)}\left(f+\frac{l}{\tau}\right) I_m\right] \tag{A9}$$

$$Q = \mathrm{e}^{i2\pi fT}\sum_{n=0}^{\infty}\left(\frac{i}{2\pi f}\right)^{n} L^n x_j(\mathrm{x}_0, T) - \sum_{n=0}^{\infty}\left(\frac{i}{2\pi f}\right)^{n} L^n x_{j0}. \tag{A8}$$

From (A8) we get

$$|Q| \le \sum_{n=0}^{\infty}\left(\frac{1}{|2\pi f|^n}\right)(|L^n x_j(\mathrm{x}_0, T)| + |L^n x_{j0}|). \tag{A9}$$

From the formal solution (A3) it is easy to see that

$$L^n x_j(\mathrm{x}_0, T) = \frac{\mathrm{d}^n}{\mathrm{d}t^n} x_j(\mathrm{x}_0, t)|_{t=T} \quad \text{and} \quad L^n x_{j0} = \frac{\mathrm{d}^n}{\mathrm{d}t^n} x_j(\mathrm{x}_0, t)|_{t=0}. \tag{A10}$$

Assuming that the trajectory $x_j(\mathrm{x}_0, t)$ of the dynamical system (A1) belongs to C^∞ class, it follows that there exists an upper bound M_n for $|\mathrm{d}^n/\mathrm{d}t^n x_j(\mathrm{x}_0, t)|, n \ge 1$. Hence we obtain

$$|L^n x_j(\mathrm{x}_0, t)| \le M_n \quad \text{for all} \quad t \in [0, \infty). \tag{A11}$$

We define a parameter $M = \sup_n (M_n)^{(1/n)}$. It can be seen that

$$\sum_{n=1}^{\infty}\frac{|L^n x_j(\mathrm{x}_0, t)|}{|2\pi f|^n} \le \sum_{n=1}^{\infty}\left(\frac{M}{|2\pi f|}\right)^n \tag{A12}$$

for all $t \in [0, \infty)$. The above series is convergent for $|2\pi f| > M(= 2\pi f_c)$. The series can then be summed to get

$$|Q| \le 2\left[M_0 + \frac{M/|2\pi f|}{1 - M/|2\pi f|}\right]. \tag{A13}$$

This illustrates that for $|2\pi f| > M(= 2\pi f_c)$, the power spectrum will have a tail

$$P_{j,T}(f) \sim \frac{M_0^2}{(\pi f)^2 T}. \tag{A14}$$

This will manifest as a $(\pi\tau/\sin\pi f\tau)^2$ tail for a discrete time series sampled at a time step τ (due to aliasing), consistent with our numerical results for the power spectra of chaotic dynamical systems.

Appendix II

Let $x(t)$ be a stationary stochastic process with an absolutely continuous spectral measure. Let $\{x_j = x(t = j\tau);\ j = 0, \ldots, N-1\}$ be a discretely sampled data set. $P(f, N, \tau)$ can be written as

$$P(f, N, \tau) = \frac{\tau}{N}\sum_{j=0}^{N-1}\sum_{k=0}^{N-1}\langle x_j x_k\rangle \mathrm{e}^{i2\pi f\tau(j-k)}. \tag{A1}$$

Since the process is stationary one can write $\langle x_j x_k \rangle = c(\tau(j-k))$. Substituting this and simplifying, we get

$$P(f,N,\tau) = \int_{-\infty}^{\infty} \mathrm{d}f' C(f') S_N\,, \tag{A2}$$

where

$$S_N = \tau\left[1 + 2\sum_{l=1}^{N-1}\left(1-\frac{l}{N}\right)\cos 2\pi\tau(f-f')l\right] \tag{A3}$$

and $C(f)$ is the true (unaliased) power spectrum. $P(f, N, \tau)$ can be equivalently represented as (see eq. (22))

$$P(f,N,\tau) = \sum_{m=-\infty}^{\infty} C\left(f+\frac{m}{\tau}\right) + R(f,N,\tau). \tag{A4}$$

Thus it is only the true spectral density $C(f)$ that goes as input to the analysis. By definition, a generic stationary stochastic process is characterized by a spectral measure which is absolutely continuous, and is nonzero for all frequencies. It is clear from (A4) that if the true power spectrum is nonzero, so is the aliased power spectrum. This implies that for a given $N, P(f,N,\tau)$ of a generic stationary stochastic process consists of both an N-independent term and terms depending on N. In the following analysis we calculate the N dependent terms explicitly to understand which term dominates in the large N limit.

Let $z = 2\pi(f-f')\tau$, then

$$S_N = \tau\left[1 + 2\mathrm{Re}\,\frac{\mathrm{e}^{iNz}-1}{\mathrm{e}^{iz}-1} - \frac{2}{N}\frac{\partial}{\partial z}\mathrm{Im}\,\frac{\mathrm{e}^{iNz}-1}{\mathrm{e}^{iz}-1}\right]. \tag{A5}$$

This can be evaluated as

$$S_N = \frac{\tau}{N}\frac{\sin^2(Nz/2)}{\sin^2(z/2)}. \tag{A6}$$

The power spectrum now takes the form

$$P(f,N,\tau) = \int_{-\infty}^{\infty} \mathrm{d}f' C(f')\frac{\tau}{N}\frac{\sin^2 N\pi\tau(f-f')}{\sin^2 \pi\tau(f-f')}. \tag{A7}$$

Noting that S_N is periodic in $1/\tau$, the above integral can be written as

$$P(f,N,\tau) = \sum_{l=-\infty}^{\infty}\frac{1}{\pi}\int_{-(\pi/2)}^{\pi/2}\mathrm{d}\xi C\left(f+\frac{l}{\tau}+\frac{\xi}{\pi\tau}\right)\frac{\sin^2 N\xi}{N\sin^2\xi}. \tag{A8}$$

Since the spectral measure is absolutely continuous, the power spectrum $C(f)$ is smooth and one can expand $C(f)$ as a Taylor's series in $\xi/\pi\tau$, to give the power spectrum the following form

$$P(f,N,\tau) = \sum_{l=-\infty}^{\infty}\left[\sum_{m=0}^{\infty} C^{(m)}\left(f+\frac{l}{\tau}\right) I_m\right] \tag{A9}$$

where

$$I_m = \frac{1}{\pi m!}\int_{-(\pi/2)}^{\pi/2} \mathrm{d}\xi \left(\frac{\xi}{\pi\tau}\right)^m \frac{\sin^2 N\xi}{N\sin^2\xi}, \quad \text{and} \quad C^{(m)}(f) = \frac{\mathrm{d}^m}{\mathrm{d}f^m}C(f). \tag{A10}$$

It is easy to see that

$$I_0 = 1 \tag{A11}$$

$$I_{2m} \sim O\left(\frac{1}{N}\right) \quad \text{and} \quad I_{2m-1} = 0, \quad m = 1, \ldots. \tag{A12}$$

With the help of these results the power spectrum can easily be written as

$$\begin{aligned} P(f,N,\tau) = \sum_{l=-\infty}^{\infty} C\left(f+\frac{l}{\tau}\right) + \frac{1}{N}\sum_{l=-\infty}^{\infty}\Bigg[C^{(2)}\left(f+\frac{l}{\tau}\right)\left(\frac{\ln 2}{\tau^2}\right) \\ + C^{(4)}\left(f+\frac{l}{\tau}\right)\left(\frac{\ln 2 - \frac{9\zeta(3)}{2\pi^2}}{2\pi^2\tau^4}\right) + \cdots\Bigg] + O\left(\frac{1}{N^2}\right), \end{aligned} \tag{A13}$$

where $\zeta(n)$ is the Riemann zeta function of order n.

Since the first term (the N independent part) in the above equation is nonzero for any absolutely continuous nonzero spectral measure and since the coefficient of $1/N$ is finite, we can always find an N such that the first term dominates. This implies that for a generic stationary stochastic process an N independent term will dominate the power spectrum at all frequencies, if N is sufficiently large.

In the case of deterministic dynamical systems, it is known that the power spectra of regular deterministic dynamical systems are discrete, and they do not have N-independent part at high frequencies. Only chaotic systems show a semblance of a continuous power spectrum. Numerical results show that chaotic systems also do not have an N-independent part at high frequencies. Thus it is not possible that the power spectra of deterministic dynamical systems are dominated by an N-independent part at high frequency region.

References

[1] See e.g., Hao Bai-Lin, *Chaos* (World Scientific, Singapore 1984)

[2] N H Packard, J A Crutchfield, J D Farmer and R S Shaw, *Phys. Rev. Lett.* **45**, 712 (1980)
D Farmer, E Ott, and A Yorke, *Physica* **D7**, 153 (1980)
J P Eckmann and D Ruelle, *Rev. Mod. Phys.* **57**, 617 (1985)
H D I Aberbanel, R Brown, J J Sidorowich, and L S Tsimring, *Rev. Mod. Phys.* **65**, 1331 (1993)
M C Valsakumar, K P N Murthy and S Venkadesan, Time Series Analysis – A Review, in *Non linear Phenomena in Materials Science III – Instabilities and Patterning* edited by G Ananthakrishna, L P Kubin and G Martin, Solid State Phenomena (Scietec Publ., 1995) vol. 42–43

[3] D E Sigeti, *Physica* **D82**, 136 (1995)

[4] D E Sigeti, and W Horsthemke, *Phys. Rev.* **A35**, 2276 (1987)

[5] U Frisch, and R Morf, *Phys. Rev.* **A23**, 2673 (1981)

[6] H Bercovici, P Constantin, C Foias, and O P Manley, *J. Stat. Phys.* **80**, 579 (1995)

[7] O E Rossler, *Phys. Lett.* **A57**, 397 (1976)

3] A Nath and D Shankar Ray, *Phys. Rev.* **A35**, 1959 (1987)
9] G Ananthakrishna and M C Valsakumar, *J. Phys.* **D15**, L171 (1982)
0] E N Lorenz, *J. Atmos. Sci.* **20**, 130 (1963)
] T Kaptaniak, *Prog. Theor. Phys.* **93**, 831 (1995)
2] Y Pomeau, and P Manneville, *Commun. Math. Phys.* **74**, 189 (1980)
3] See e.g., N C Geckinli, and D Yavuz, *Discrete Fourier transformation and its applications to power spectra estimation* (Elsevier, Amsterdam 1983)

Dynamics of simple one-dimensional maps under perturbation

SOMDATTA SINHA and PARICHAY K DAS*
Centre for Cellular & Molecular Biology, Uppal Road, Hyderabad 500 007, India
*Indian Institute of Chemical Technology, Uppal Road, Hyderabad 500 007, India

Abstract. It is known that the one-dimensional discrete maps having single-humped nonlinear functions with the same order of maximum belong to a single class that shows the universal behaviour of a cascade of period-doubling bifurcations from stability to chaos with the change of parameters. This paper concerns studies of the dynamics exhibited by some of these simple one-dimensional maps under constant perturbations. We show that the "universality" in their dynamics breaks down under constant perturbations with the logistic map showing different dynamics compared to the other maps. Thus these maps can be classified into two types with respect to their response to constant perturbations. Unidimensional discrete maps are interchangeably used as models for specific processes in many disciplines due to the similarity in their dynamics. These results prove that the differences in their behaviour under perturbations need to be taken into consideration before using them for modelling any real process.

Keywords. Bifurcation; chaos; ecology; one-dimensional map; perturbation.

PACS Nos 05.45; 80.89.60; 03.20

1. Introduction

The field of dynamical systems and, especially the study of chaotic systems, has been hailed as one of the important breakthroughs in science in this century. Past three decades have seen an explosion of interest in the study of nonlinear dynamical systems in almost all disciplines, and chaotic and random behaviour of solutions of deterministic systems is now understood to be an inherent feature of many nonlinear systems. This has been possible due to infusion of interest and techniques from a variety of fields, from physics and chemistry to biology and social sciences as most natural systems seem to be complex and nonlinear. With renewed interest in low dimensional discrete dynamical systems in physics and other disciplines such as mathematical ecology and economics, it was shown that even very simple mathematical models can actually exhibit very complicated dynamics. In his seminal paper, May [1] urged that simple one-dimensional maps should be part of early mathematical education since on one hand it is simple to analyse, yet on the other hand, the range of complex dynamics exhibited by it can introduce the concepts of variability, chance and imbalance as perfectly natural behaviour expected from natural systems.

The general form of a one-dimensional discrete equation (map) is

$$X_{t+1} = f(X_t), \tag{1}$$

where X_{t+1} is a function of the discrete variable X at the earlier (i.e. tth) iteration and some parameters. Historically [1, 2], X_t has been studied as the population density of organisms (whose generations do not overlap) at the tth generation. If the function $f(X_t)$ is such that it has the propensity to increase monotonically when X is small and to decrease monotonically when it is large due to density-dependent processes, then it assumes a "hump"-like shape with a unique maximum.

More than two decades ago [1, 3–5] the numerous discrete models with single-hump growth functions existing in the ecological literature were placed under a general class of unidimensional maps based on the similarity in their dynamic behaviour with increasing growth rates. It was shown that all these maps display a "universal" bifurcation structure where, with the increase in a parameter (e.g., the growth rate), the system dynamics changes from stable equilibrium point to chaos through the period-doubling route, and the chaotic regime contains an ordered sequence of periodic windows [1, 2, 6]. This similarity in bifurcation structure of these one-dimensional maps initiated the use of these models interchangeably to describe experimental population data in ecology [7–12].

Though it is known that all one-hump maps having the same order of maximum belong to the same universality class [13], the precise conditions that determine if a single-hump nonlinear function would belong to the "universal" class are not clear [14]. Recent studies have shown that slight perturbations of a map can cause breakdown or distortion of universal behaviour [14–18]. These studies, from mathematical ecology again, describe the different responses of two maps (logistic and exponential) belonging to the same universality class when an additive disturbance (can be termed as perturbation or bias) is imposed on the nonlinear functions.

In this paper we study a few representative maps that are used commonly in theoretical studies and also to model real systems [1, 19, 20] under such perturbation and show the similarities and differences of their response. The major results are:

(a) the logistic map shows very different response to perturbations compared to other maps of the same universality class which have long 'tail' or 'plateaus' at high X; and

(b) the response dynamics depends on the details of the shape of the hump – positive perturbation enhances the effect of the 'tail', whereas, negative perturbation exposes the gradient of the upper part of the hump. Thus, the logistic map stands out among other maps whose 'plateaus' and asymmetric shapes give rise to a variety of interesting, unexpected and unusual dynamics. Since these latter type of maps are considered to be more realistic, one may anticipate many unexpected dynamics in nature.

The paper is organized as follows:

(i) Three commonly used one-dimensional maps and the types of perturbations used for this study are introduced.

(ii) The effect of the perturbations on the dynamics of these maps is shown.

(iii) The results are discussed with a view to delineate the differences in these maps.

(i) *One-dimensional maps*

We choose the following three one-dimensional equations with the specific single hump

Table 1. Properties of one dimensional single hump maps.

Maps	x-Domain	r-Range	Equilibrium points $f(x)=x$	Critical point $f'(x)=0$
Logistic map	$0 \le x \le 1$	$1 < r \le 4$	0 and $(r-1)/r$	$1/2$
Exponential map	$0 \le x < \infty$	$0 < r$	0 and 1	$1/r$
Bellows map	$0 \le x < \infty$	$1 < r$	0 and $(r-1)^{1/b}$	$\left(\frac{1}{b-1}\right)^{1/b}$

functions for our study:

(I) Logistic model:

$$X_{t+1} = rX_t(1 - X_t). \tag{2a}$$

(II) Exponential model:

$$X_{t+1} = X_t \exp[r(1 - X_t)]. \tag{2b}$$

(III) Bellows model:

$$X_{t+1} = \frac{rX_t}{(1 + X_t^b)}\,. \tag{2c}$$

Here 'r' and 'b' are parameters. These three models were chosen because they are most commonly used, both theoretically and experimentally [1, 3–5, 11, 12, 14, 19–22]. Though eq. (2a) is the simplest among the non-linear difference equation models, the other two equations (2b), (2c) have more secure provenance in the biological literature [1, 12, 14, 21]. Table 1 summarizes the detailed features of these models, and figures (1a–f) show the return maps (X_t versus X_{t+1} plot) and the bifurcation diagrams with increasing r for each model. Figure 1 clearly describes the similar "inverted U-shape" (hump) nature of $f(X)$–the nonlinear function–in the return maps, and the commonality in their dynamics of progression from stable point, to a hierarchy of stable cycles of period 2^n, to chaos with increasing r. These properties are observed in all models from this class of unidimensional single-hump equations [1, 2, 5].

Constant perturbation. Under an additive constant perturbation the one-dimensional equation (1) takes the form

$$X_{t+1} = f(X_t) + L = F(X_t) \tag{3}$$

where L is a parameter which can take both positive and negative values. The effect of a fixed positive and negative L on $f(X)$ is to move the humps shown in the return maps in figures (1a, 1c and 1e) above or below the X-axis respectively. As examples this is shown for the logistic model for $r = 3.9$ in figure 2a, and for the Bellows model for $r = 5$ and $b = 6$ in figure 2b. The continuous curves represent the unperturbed return maps (eqs (2a) and (2c)) and the dashed and dotted curves are for the perturbed maps ($+L$ and $-L$ respectively) in eq. (3). The exponential model is similar to the Bellows model.

It is clear from figure 2 that the perturbations change the intersection of the return maps (dashed curve and dotted curve) with the $X_{t+1} = X_t$ line (solid line), thus changing the

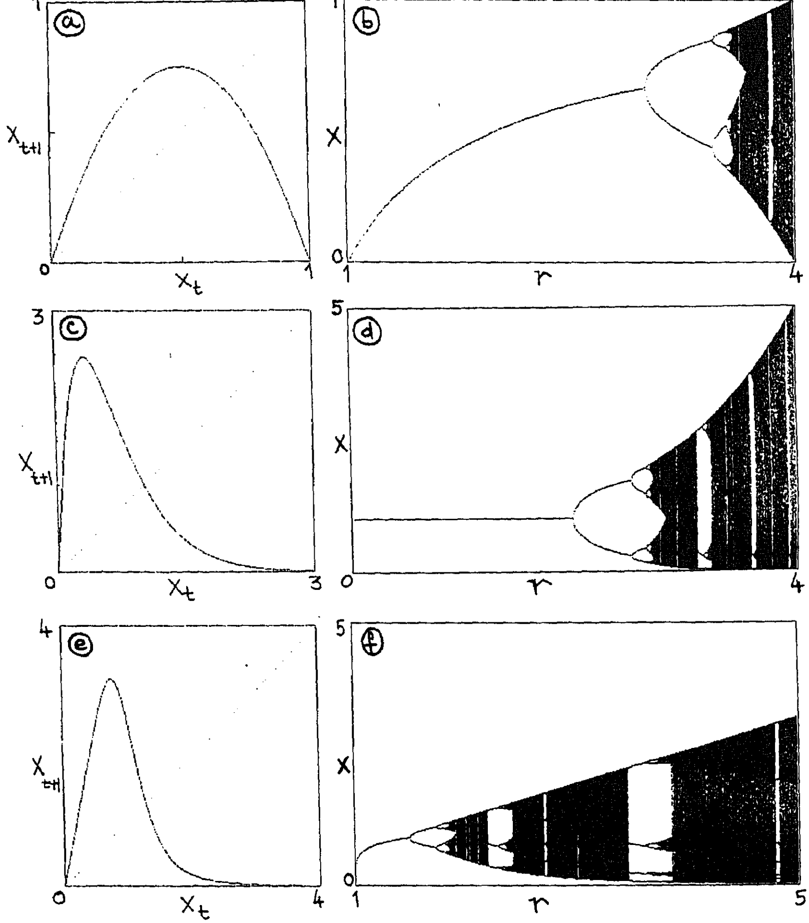


Figure 1. Return maps and bifurcation diagrams, respectively, for the (**a**, **b**) logistic, (**c**, **d**) exponential, and (**e**, **f**) Bellows maps. Unless otherwise stated the initial conditions are chosen to be X at which $f'(X) = 0$ in all figures.

number of positive equilibrium points and their stability. With negative L, there are two equilibrium points (x_1, x_2 in figure 2), and the first equilibrium point (x_1) is unstable leading all iterations below that critical value to zero. The dynamics exhibited by these maps depend on the gradient of the hump at the second equilibrium point (x_2). The other major change that negative perturbation introduces in the exponential and Bellows maps is that these two maps are no longer valid for all $X > 0$, thus introducing limits in both r and X within which the iterations will remain positive. In the case of the exponential and the Bellows maps (see figure 2b), positive L creates a higher 'floor' as the tail of the maps are elevated. This does not happen in the case of the logistic map and it continues to be valid only within a range of X. In this paper whenever X assumes zero or negative values it is termed as 'escape'.

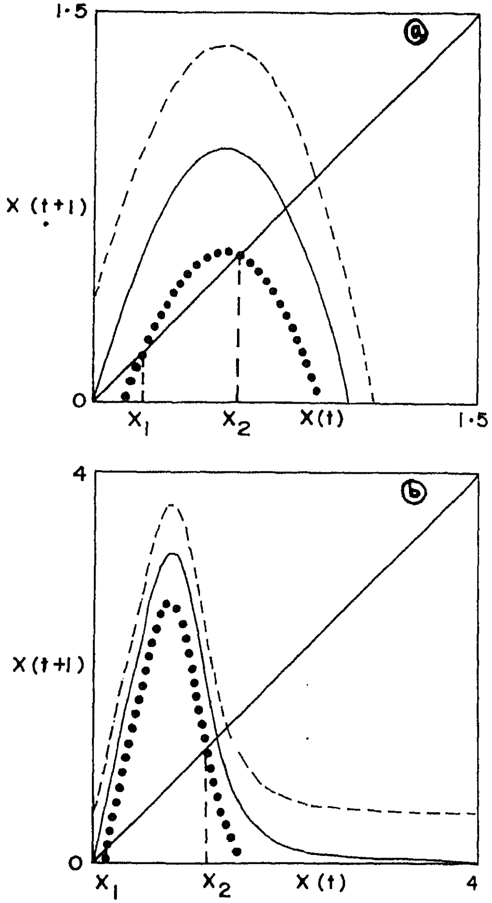


Figure 2. Return maps of the (**a**) logistic, and (**b**) Bellows maps. Continuous curve – unperturbed map, Dashed curve – maps with positive perturbation (+*L*). Dotted curve – maps with negative perturbation (−*L*).

(ii) *The dynamics of the perturbed maps*

The dynamics exhibited by the three models with positive and negative L are shown through bifurcation and stability diagrams in the $r-L$ parameter space ($r-L$ diagrams). Interestingly whatever be the magnitude of L (except when it pushes the system to "escape"), the dynamical system always returns to its unperturbed state on withdrawal of the perturbation. Therefore, we have studied the response of the maps to large ranges of L values. Figures 3, 4 and 5 are the bifurcation diagrams showing the dynamics of the perturbed maps with increasing L at two r values corresponding to stable and chaotic dynamics at the unperturbed state.

Logistic map. Figures 3a–d show the dynamics exhibited by the logistic map with increasing positive and negative perturbations for stable r ($r = 2.0$) and for chaotic r ($r = 3.9$) respectively. It is clear that positive perturbation induces (figure 3a) or

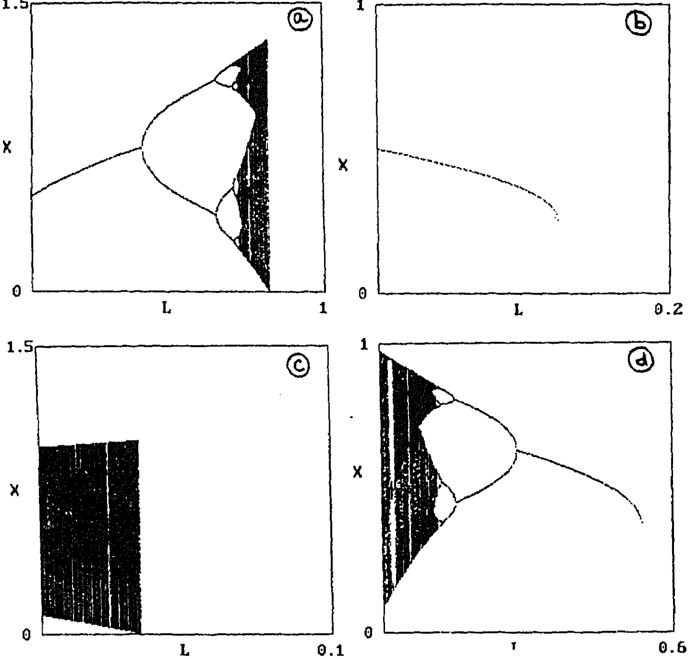


Figure 3. Bifurcation diagrams with L for the logistic map for stable $r = 2.0$. (**a**) positive L, (**b**) negative; and, L chaotic $r = 3.9$ (**c**) positive L, (**d**) negative L.

maintains (figure 3c) chaotic dynamics for all r, though 'escape' occurs for very small perturbations in a chaotic logistic map. On the other hand, negative perturbation induces stability in a chaotic map (figure 3d), though 'escape' ensues for stable r at relatively low negative L (figure 3b). Thus the final response of the logistic map to both types of perturbations is opposite and also independent of r.

Exponential map. Figures 4a–d show the effects of increasing positive and negative perturbations in the exponential map for stable ($r = 1.8$) and chaotic r ($r = 3$) respectively. In contrast to the logistic map (figure 3a), a stable exponential map remains stable under both types of perturbations (figures 4a and 4b), and a small amount of positive perturbation can induce stability in a chaotic map also (figure 4c). The response of the chaotic map to negative perturbation is however unusual. Though, like the logistic map in this case also increasing negative perturbation shows overall period reversal to stability from chaos, yet this map shows an intervening band of L values at which 'escape' occurs, but positive iterates recur for a range of higher L beyond which iterates 'escape' for all L. Interestingly, though in this map both types of perturbations finally render the map stable for all r, 'escape' is observed with $-L$ only for parameter (r) values corresponding to higher period ($P \geq 4$) dynamics.

Bellows map. It is clear from figures 5a–d that the response of the bellows map is very similar to the exponential map for stable r ($r = 1.2$) under negative L, and for chaotic r

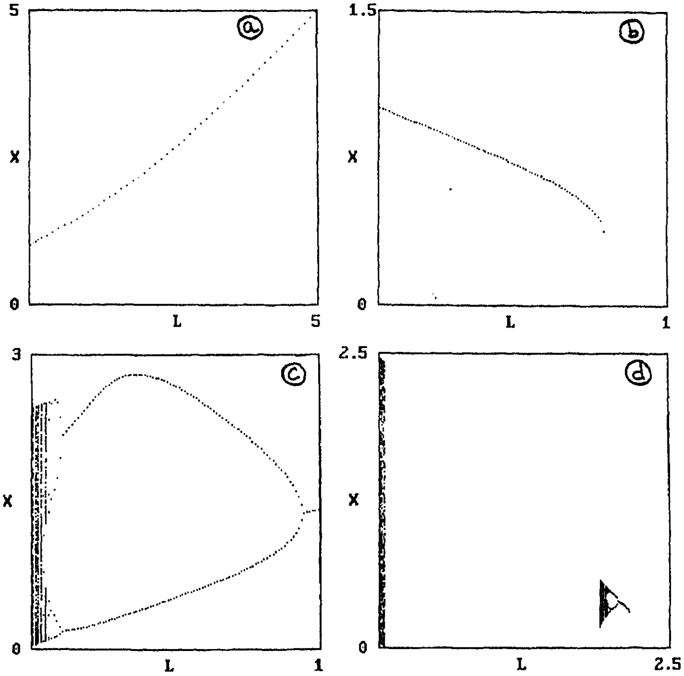


Figure 4. Bifurcation diagrams with L for the exponential map for stable $r = 1.8$. (**a**) positive L, (**b**) negative; and, L chaotic $r = 3.0$ (**c**) positive L, (**d**) negative L.

($r = 3.0$) under both positive and negative L. The stable-unstable-stable 'bubble' in figure 5a with increasing positive perturbation in a stable map is a reminder of the logistic map (see figure 3a) except that further period doubling bifurcations are curtailed here and period reversal gives rise to stable dynamics at higher L. It is also interesting to note that, like the logistic map (figure 3b), stable bellows map (figure 5b) also shows 'escape' for small negative perturbation. This map has an additional parameter 'b' which has a role in deciding the shape of the hump and hence the bifurcation scenario. We have taken $b = 6$ here which imposes strong nonlinearity in the maps, though similar behaviour is observed for lower b (> 2) also.

To get a complete picture of the effect of the constant perturbations on the dynamics of the maps for a range of values of the nonlinear parameter r, we have used a combined analytical and numerical approach to obtain the stability regions in the $r - L$ parameter space. Figure 6 shows three $r - L$ diagrams which provide a description of the different dynamical behaviour such as, stable fixed point ($P = 1$), stable oscillations of period 2 ($P = 2$), and complex oscillations of higher periods including chaos ($P \geq 4$) that are exhibited by the models for different values of r and L. The space outside the demarcated region constitutes the r and L values that lead to 'escape'. Since positive and negative L moves the hump vertically up and down, the range of X and r values for which the logistic map is valid also changes with L. For negative L, the exponential and Bellows maps both resemble the logistic map since they now do not have the extended tail and hence exist

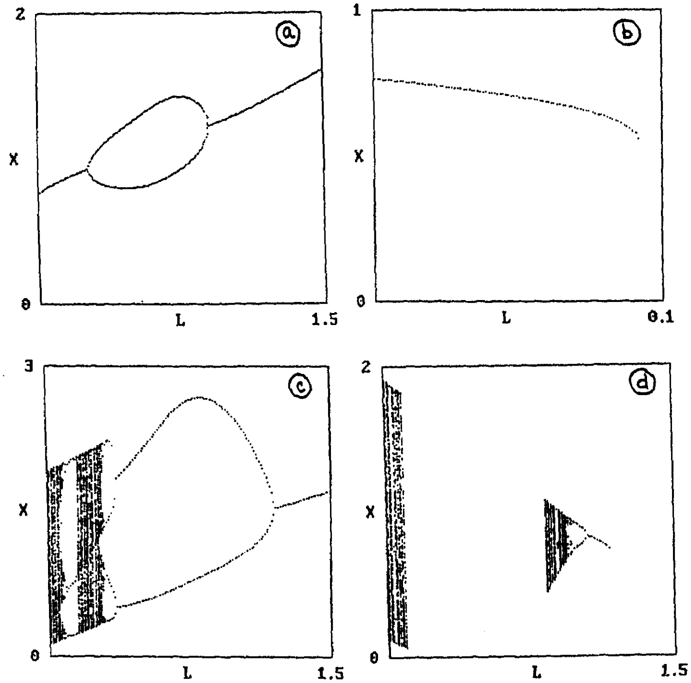


Figure 5. Bifurcation diagrams with L for the bellows map for stable $r = 1.2$. (**a**) positive L; and, (**b**) negative L; and, chaotic $r = 3.0$ (**c**) positive L, (**d**) negative L.

only within a certain range of r. Therefore they also show the 'escape' region for higher negative perturbations.

The boundaries in the $r - L$ plot in figure 6 can be obtained as follows:

(a) The curve that forms the lower bound for all maps in the $r - L$ plot for negative L is obtained by finding the values of L for every r at which a saddle-node bifurcation takes place [20].

(b) The intervening 'escape' region observed in the exponential and bellows maps at higher r occurs for intermediate values of L at which, $F(X_{\max}) - x_1 < 0$ where x_1 is the first unstable equilibrium point (cf. figure 2b). Thus, the boundary in the $r - L$ parameter space can be obtained by considering values of r and L which satisfy the expression

$$F(X_{\max}) - x_1 = 0.$$

This condition is never satisfied in the case of the logistic map.

(c) The uppermost curve for $+L$ separating the 'escape' region is observed only for the logistic map (figure 6a). In the logistic map, there is one L for every r at which $F^2(X_{\max}) \leq 0$, i.e. the iterate falls outside the hump on the X-axis. Both exponential and Bellows maps do not have this region since they have 'tails' for high X (for $+L$) which force the iterates to be positive.

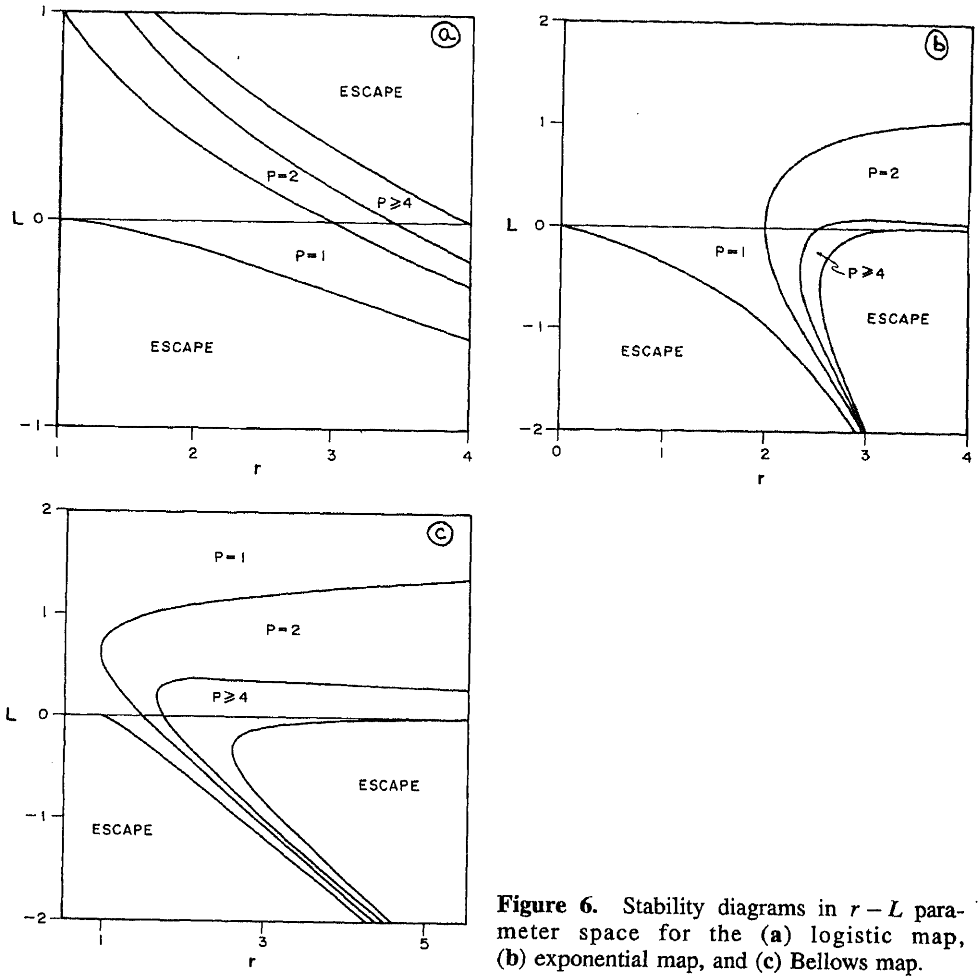


Figure 6. Stability diagrams in $r - L$ parameter space for the (**a**) logistic map, (**b**) exponential map, and (**c**) Bellows map.

(d) The boundaries of the first two period doubling bifurcations ($P = 2$ and $P = 4$) in figures 6a–c are obtained by solving the equations

$$F^n(X) = X \quad \text{and} \quad [F^n(X)]' = -1, \quad n = 1, 2.$$

Figure 6 clearly shows that the response of the logistic map to perturbations is very different when compared to that of the exponential and the Bellows maps. This feature clearly indicates that all these maps can not be treated at par.

Though from above the exponential and the Bellows maps seem to have very similar response to both positive and negative perturbations, comparison of figures 4a and 5a indicate finer differences between them. The ‘bubble’ of instability in the bellows map (figure 5a) is a feature which arises due to the detailed shape of $f(X)$ near the critical point. Figure 7 exemplifies this for the three maps. In this figure the continuous curve shows the slope of the map functions $f'(X)$ for all X starting after the critical point for the r value at which the first bifurcation from stability to oscillation occurs. The dotted curve represents the quantity $\{f(X) - X\}$.

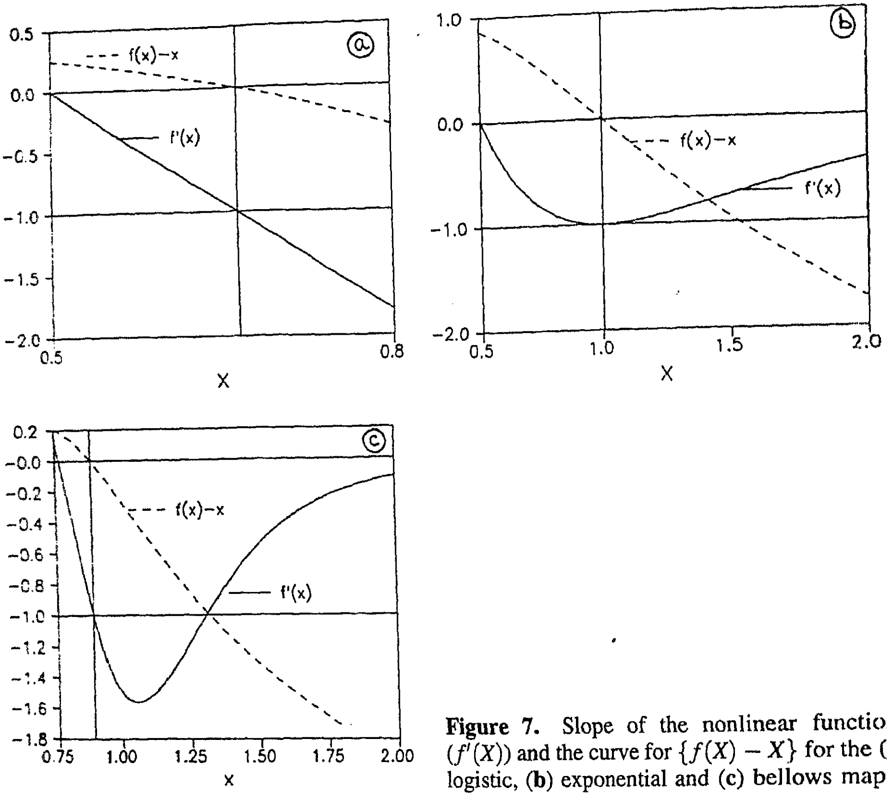


Figure 7. Slope of the nonlinear functions ($f'(X)$) and the curve for $\{f(X) - X\}$ for the (**a**) logistic, (**b**) exponential and (**c**) bellows maps.

It is clear from these figures that the equilibrium point x_2 (cf. figure 2b) where $\{f(X) - X\} = 0$ occurs is also the point at which $f'(X) = -1$. The major point to be noted is that there are two values of X at which $f'(X) = -1$ for the Bellows map in figure 7c as opposed to the exponential map in figure 7b. Here (figure 7c), the equilibrium point is closer to the critical point of the map when the first bifurcation (at $r = 1.5$) occurs. A lower value of r will obviously show stable dynamics in the unperturbed map and the equilibrium point will be on the left of the first intersection of $f'(X) = -1$. A constant positive perturbation (i.e. $+L$) to $f(X)$ at stable r will move the equilibrium point to the right on the X-axis. The system will then exhibit stable-unstable-stable 'bubble' as L is increased due to the presence of the two intersections of $f'(X) = -1$ in the Bellows map. Thus it is clear from this figure that there exist a range of stable r values in this map where positive perturbations can give rise to "bubble" like dynamics. Such a situation never arises for both the logistic and the exponential maps due to the shape of their humps (see figures 7a, b).

2. Discussion

The above-mentioned results clearly demonstrate that small perturbations can delineate the differences among the maps which are considered to belong to the same universality class and have similar bifurcation dynamics. They can now be classified based on their response to constant perturbations which is model-specific. The main features that play important roles in the manifestation of the differences in dynamics in these maps under perturbations are

1. The presence of the 'tails' in the nonlinear functions in the exponential and Bellows maps at higher X. This gives rise to higher 'floors' (plateaus which flattens out at high X) with positive perturbations, leading to period reversals to stability [14].
2. The detail shapes of the nonlinear functions near the maximum at medium values of X. These are highlighted in the form of 'bubbling' under positive perturbations, or, the minimum iterate $(F(X_{\max}))$ falling below the first unstable equilibrium point under negative perturbations.
3. The symmetric, purely concave-down structure of the hump in the logistic model does not allow any of the above processes to occur under constant additive perturbations. In fact the nonlinear parameter r in the logistic map can be easily scaled to include the effect of the perturbation (L). This can explain the changes in the dynamics with positive and negative perturbation in the logistic map in terms of increased or decreased effective r.

These simple one-dimensional discrete equations studied here have been extensively used for describing density-dependent population growth of organisms having non-overlapping generations [3, 5, 9, 23, 24]. In ecological terms, the constant perturbation can be considered as constant migration to/from a population. We have shown that these maps do not behave similarly under perturbations thus exposing the inherent non-equivalence among them. Our results have far-reaching implications not only in the study of one-dimensional non-linear discrete systems in general but also in the area of population dynamics and resource management studies. Thus one can say that though all unidimensional, discrete, single-hump maps show similar bifurcation dynamics [1, 2, 5], the exact nature of the non-linear functions plays an important role in their different dynamic response to simple perturbations.

Acknowledgement

This work is a part of the project funded by Department of Science and Technology, India. We gratefully acknowledge its support.

References

[1] R M May, *Nature* **261**, 459 (1976)
[2] M J Feigenbaum, *J. Stat. Phys.* **19**, 25 (1978)
[3] R M May, *Science* **186**, 645 (1974)
[4] R M May, *J. Theor. Biol.* **51**, 511 (1975)
[5] R M May and G F Oster, *Am. Natur.* **110**, 573 (1976)
[6] N Metropolis, M L Stein and P R Stein, *J. Combin. Theory* **15**, 25 (1973)
[7] C J Krebs, *Ecology: The experimental analysis of distribution and abundance*, 2nd edn (Harper International Edition, Harper and Row, New York, 1978)
[8] J A Yorke, N Nathanson, G Pianigiani and J Martin, *Am. J. Epidem.* **109**, 103 (1979)
[9] C W Clark, *Bioeconomic modeling and fisheries management* (Wiley, New York, 1985)

[10] J D Murray, *Mathematical Biology* (Springer-Verlag, Berlin, 1989)
[11] N V Joshi and M Gadgil, *Theor. Pop. Biol.* **40**, 211 (1991)
[12] H I McCallum, *J. Theor. Biol.* **154**, 277 (1992)
[13] G Ambika and K Babu Joseph, *Pramana – J. Phys.* **39**, 193 (1992)
[14] L Stone, *Nature* **365**, 617 (1993)
[15] R D Holt, in *Population biology – Lecture notes in biomathematics* edited by H I Freedma and C Strobeck (Springer Verlag, Berlin, 1983) vol. 52
[16] S Sinha and S Parthasarathy, *J. Bios.* **19**, 247 (1994)
[17] S Parthasarathy and S Sinha, *Phys. Rev.* **E51** 6239 (1995)
[18] S Sinha and S Parthasarathy, *Proc. Natl. Acad. Sci. (USA)* **93**, 1504 (1996)
[19] T S Bellows, *J. Anim. Ecol.* **50**, 139 (1981)
[20] R L Devaney, *An introduction to chaotic dynamical systems* (Menlo Park, Ca: Th Benjamin/Cummings Pub. Co. Inc., 1986)
[21] M P Hassell, J H Lawton and R M May, *J. Anim. Ecol.* **45**, 471 (1976)
[22] K Masutani, *Bull. Math. Biol.* **55**, 1 (1993)
[23] J R Beddington and R M May, *Science* **197**, 463 (1977)
[24] W M Schaffer and M Kot, *J. Theor. Biol.* **112**, 403 (1985)

Chaos in modulated logistic systems

V M NANDAKUMARAN
International School of Photonics, Cochin University of Science and Technology, Cochin 682 022, India

Abstract. This paper is a review of the work done on the dynamics of modulated logistic systems. Three different problems are treated, viz, the modulated logistic map, the parametrically perturbed logistic map and the combination map obtained by combining two maps of the quadratic family. Many of the interesting features displayed by these systems are discussed.

Keywords. One dimensional maps; universality; parametric perturbation; scaling relations.

PACS Nos **05.40; 05.45; 47.52**

1. Introduction

The complicated dynamical behaviour that can arise in the low dimensional deterministic nonlinear systems has generated a lot of intense research in the last two decades [1, 2]. These studies have shown that simple physical systems obeying simple laws can often show complex and chaotic behaviour. Since several nonlinear systems can be analysed using one dimensional maps, such maps have come to play an important role in understanding many of the general features of chaotic systems. One of the simplest one dimensional maps, namely the logistic map, originally introduced in the study of the population dynamics of non overlapping generations in biology [3] has become a paradigm for such systems. The logistic map is a one-hump map of the form

$$X_{n+1} = 4\lambda X_n(1 - X_n) \tag{1}$$

defined in the interval (0, 1) and is characterized by a control parameter λ which is also varied in the interval (0, 1). This system undergoes a transition to chaos via the well-known period doubling route and shows universal metric [4] as well as structural properties [5].

One of the earliest experimental observations of the period doubling route to chaos was in CO_2 laser [6]. However, in many of the experimental situations the control parameter can have a time dependence of its own. Therefore, it would be interesting to consider the dynamics of a map with a time-dependent control parameter.

In one of the earlier studies, Kapral and Mandel [7] investigated a non autonomous quadratic map wherein the control parameter was assumed to vary linearly with time. They have observed that this time dependence of the parameter delayed the onset of bifurcations in the system. The system also showed bistability and hysteresis. Ruelle had suggested the study of dynamical systems with adiabatically fluctuating parameters where the control parameter has a very slow variation in time and this time dependence itself might be determined by a dynamical system [8].

In this paper the dynamics of the modulated logistic system is presented. We use the term modulated logistic system (MLS) to describe the following three situations:

(i) A logistic map in which the control parameter has a discrete time-dependence determined by a dynamical system – a modulated logistic map (MLM).
(ii) A logistic map wherein the control parameter is perturbed by the addition of a periodic term (parametrically perturbed logistic map).
(iii) A combination map where the logistic map is combined with another map of the same quadratic family.

The paper is organized as folllows. In § 2, we introduce the MLM and present some of the earlier results of Harikrishnan and Nandakumaran [9–12]. Some of the related works in this field are also discussed in the section [13]. We assume that the time evolution of the control parameter is determined by another logistic map. This coupled system shows many interesting features such as the locking of the periodicities of the two subsystems. Section 3 deals with the parametrically perturbed logistic map. Here the control parameter of the logistic map is perturbed by a sequence of periodic pulses. Such a system is relevant for modelling those systems which are subjected to periodic stimuli. The dynamics of the system shows many interesting and novel features such as the formation of bubble structure in the bifurcation diagram, emergence of periodic behaviour after a chaotic regime etc. The combination map is presented in § 4. This is a one hump map obtained by adding a sinusoidal map to the logistic map (both the maps belong to the same universality class). It is found that the scaling index characterizing the behaviour of the Lyapunov exponent near the onset of the chaotic transition is different from that of a single one-hump map. Section 5 contains the concluding remarks.

2. Modulated logistic map

In this section we briefly review the results obtained by Harikrishnan and Nandakumaran on MLM [9–12]. An MLM is defined by the following pair of equations:

$$X_{n+1} = 4\lambda_n X_n(1 - X_n), \tag{2}$$

$$\lambda_{n+1} = 4\mu\,\lambda_n\,(1 - \lambda_n) \tag{3}$$

with $0 < X_n, \lambda_n, \mu < 1$. Here μ plays the role of the control parameter. This system is analogous to the one suggested by Ruelle although in the present case the time evolution of the λ-system is not slow, but is of the same order as that of the X-system. As μ is varied continuously, the MLM undergoes a sequence of period doubling bifurcations. However, the bifurcation diagram for X is very much different from the one for the λ-system [11].

In order to establish the universality properties of the map we have determined the Feigenbaum constant δ using the method of eigenvalue matching renormalization due to Derrida *et al* [14, 15]. The basic idea of the method is the following. Let us represent the

MLM as

$$(X_{n+1}, \lambda_{n+1}) = T_\mu(X_n, \lambda_n). \tag{4}$$

We linearize the map $T_\mu^{(n)}$ in the neighbourhood of an n-cycle. With each μ we associate a μ' such that the linearization of $T_\mu^{(n)}$ around a point of cycle n is identical to linearization of $T_{\mu'}^{(2n)}$ around a point of cycle $2n$. This gives rise to a relation connecting μ and μ', the fixed point of which is given by μ_∞ universal constant δ is then given by

$$\delta = \left.\frac{\mathrm{d}\mu}{\mathrm{d}\mu'}\right|_{\mu}. \tag{5}$$

The details of the calculations are presented in [10]. In the first order approximation, we choose $n = 1$, this gives a value of $\delta = 4.4339...$ which is consistent with the Feigenbaum constant for the logistic map. Better agreement can be obtained by choosing higher values for n.

One of the interesting features that we have observed in MLM is that the periodicity of X is identical to that of λ except in the range of values $0.848 < 0.860$ [9, 13] where λ has a periodicity of 2 while X has a periodicity of 4. As μ is varied λ period doubles and so does the variable X. This locking of the periodicities of the two is a special case of the more general situation described by Batra and Varma. For a given value of μ they have derived conditions under which the X-system is in a state of periodicity n or in a state of periodicity which is an integral multiple of n, when the λ-system is in a stable state of periodicity n. For the MLM the two periodicities are identical except in the range of μ mentioned above. This result has an important consequence. It shows that the MLM has a stable n-cycle for the same range of μ for which the logistic map has an n-cycle. Thus the periodicity of X is enslaved to the periodicity of λ. This implies that the ordering of the cycles in MLM is the same as that of the logistic map. Thus the Sarkovskii ordering [16] which is an universal structural property of unimodel maps is maintained in the MLM. Batra and Varma have suggested other possible 2D-systems that may show Sarkovskii ordering. These universal properties of MLM makes it suitable to model physical systems in which the control parameter is time-dependent.

3. Parametrically perturbed logistic map

In this section we consider a logistic map whose control parameter is perturbed by the addition of a periodic term [17]. Instead of changing λ continuously it is changed in a discrete sequence of pulses, repeated periodically, the envelope of the amplitudes of these pulses forming a positive sine profile. Maps where the control parameter is perturbed by a sequence of pulses may be relevant in the study of certain biological systems subjected to periodic stimuli [18]. A logistic system with a similar perturbation but with a cosine profile that includes both the positive and the negative half cycles has been considered earlier [19]. It has been shown that the map undergoes a transition from a fixed point to a state which has the periodicity of the perturbation. We observe that the map has some novel features not considered earlier. Quadrative maps with additive periodic forcing that leads to bistability and the coexistence of multiple attractors have been studied by Sanju and Varma [20].

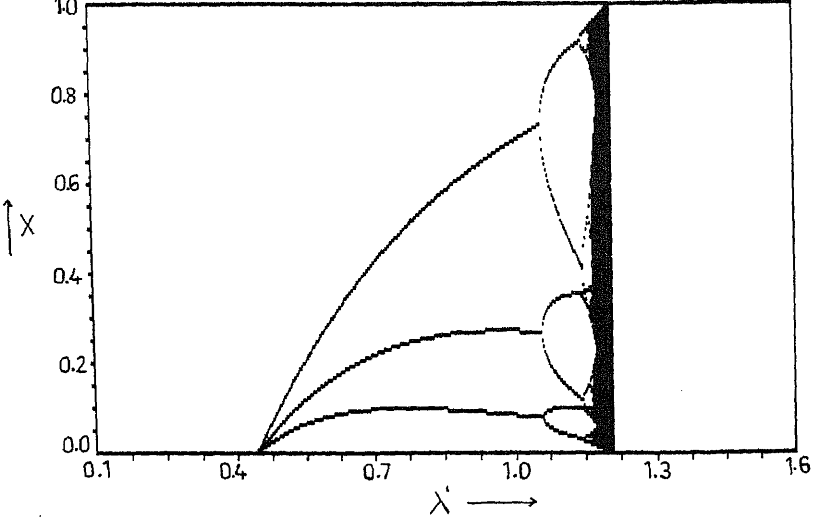


Figure 1. Bifurcation diagram of the parametrically perturbed map for $q = 3$.

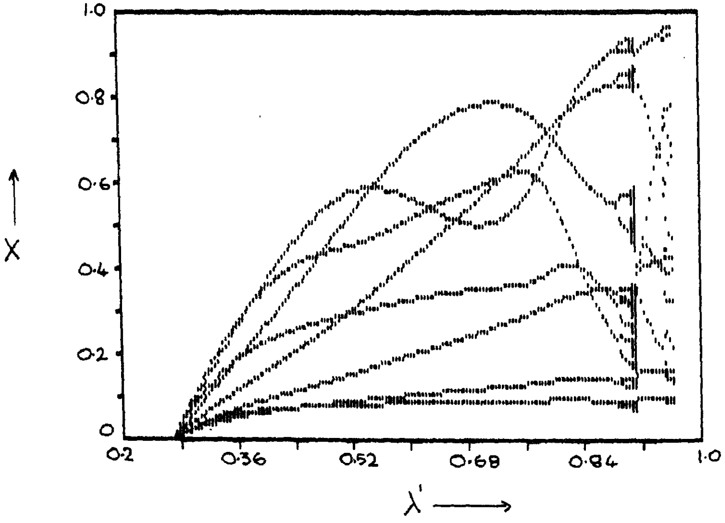


Figure 2. Bifurcation diagram for $q = 8$.

The parametrically perturbed logistic map is defined by the equations

$$X_{n+1} = 4\lambda X_n(1 - X_n),$$

with

$$\lambda = \lambda_0 + \lambda^* \sin(\pi\omega n), \quad (\omega n) \bmod 1. \tag{6}$$

This represents a train of pulses whose periodicity is determined by ω.

If $\omega = p/q$, the control parameter forms a periodic q-sequence. In what follows we chose $q = 1$. ω can also be chosen as irrational. The dynamics of the map (6) is studied

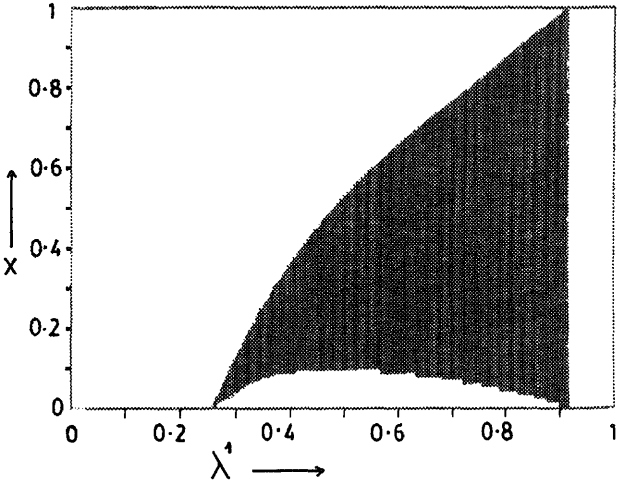


Figure 3. Bifurcation diagram for $q = (5^{1/2} - 1)/2$.

numerically by fixing λ_0 and varying the strength of the perturbation amplitude λ^*. To obtain a wide tunability for λ^* we fixed a low value for $\lambda_0 = 0.1$. For this value of λ_0, the unperturbed logistic map has a stable fixed point $X^* = 0$. The bifurcation diagrams are drawn for $q = 3, 8$ and $(5^{1/2} - 1)/2$ which is the golden mean (figures 1, 2 and 3). For $q = 3$, when λ^* is increased the fixed point $X^* = 0$ becomes unstable and the system makes a transition to a stable 3-cycle and remains locked to it for a very long range of λ^*. After this, period doublings follow in quick succession followed by a chaotic regime. When λ^* is increased still further the iterates leave the interval $(0, 1)$ and escape to infinity. Qualitatively similar features are seen for $q = 8$. However, for $q = 8$, the chaotic region is followed by the reappearance of 8-cycles which further period doubles and goes to chaos. Some of the other novel features observed in this system have not been included in this paper since they form the contents of another paper which has been sent for publication. When $q = (5^{1/2} - 1)/2$, the stable fixed point becomes unstable as λ^* is increased and goes to a quasiperiodic state (figure 3) and remains in that state until escape. When ω is chosen as irrational there are additional interesting phenomena such as strange nonchaotic attractors [21]. However, we do not consider this in the present case.

It is instructive to study the superstable cycles in this map. For superstable cycle, $x = 1/2$ is one of the cycle elements. Metropolis *et al* [5] have studied extensively the sequence of iterates starting from the extremum at $x = 1/2$. These sequences are often called the MSS sequences. The superstable cycles are represented by a sequence of symbols R and L depending on whether the iterate falls on the right (R) or the left (L) of the extremum at $x = 1/2$. We have done a similar analysis for the map (6) and the results are tabulated in table 1 for various q-values. It is clear from the table that the map (6) can support sequences that do not fall under the MSS classification. In fact the first superstable sequence for each ω contains only the symbol L. In the table the sequences indicated by a star belong to the MSS classification.

The parametrically perturbed map is of importance since it may be used to model many experimental situations such as a laser system [22] where the control parameter would be modulated by periodic pulses or pulse trains.

Table 1. Super stable sequences of the parametrically perturbed logistic map. The super stable value of λ^* is indicated by Λ^*.

q		Sequences	Λ^*
3	$k=3$	L^2	0.6672028
	$k=6$	L^2RL^2	1.008131
4	$k=4$	L^3	0.5871376
		$*RL^2$	0.6724169
		L^2R	0.9041113
5	$k=5$	L^4	0.5268982
6	$k=6$	L^5	0.4899559
		$*RL^4$	0.5539654
		L^4R	0.7129956
7	$k=7$	L^6	0.4855321
		$*RL^5$	0.59889335
		L^4R^2	0.6514045
		$*RL^4R$	0.80803725
8	$k=8$	L^7	0.4585666
		$*RL^6$	0.5080377
		L^5R^2	0.5843965
		R^3L^4	0.655407
		L^4R^3	0.81519066
		L^4RLR	0.92404916
9	$k=9$	L^8	0.4621206
		RL^7	0.4633298
		R^2L^6	0.5403486
		R^3L^5	0.7149534
		$*RL^5R^2$	0.720069367
		RL^5R^2	0.80574375
		L^4R^4	0.85825292
		$*RL^4R^3$	0.9043900
10	$k=10$	L^9	0.44201028
		$*RL^8$	0.48165978
		L^7R^2	0.508807
		R^3L^6	0.57816015
		L^5R^4	0.691275
		$*RL^5R^3$	0.7179589
		$RLRL^5R$	0.7626641
		$RLRLRL^4$	0.7773538
		L^4R^2LRL	0.8562051

*Indicates the cycles that occur in the MSS sequences. If k is the period of the cycle the sequence contains $(k-1)$ symbols.

4. Dynamics of a combination map

The combination map is obtained by combining the logistic map with a sinusoidal map. Thus we have a two-parameter one dimensional map given by [23, 24]

$$X_{n+1} = f(X_n, \lambda, A) = 4\lambda X_n(1 - X_n) - A\sin(\pi X_n). \tag{7}$$

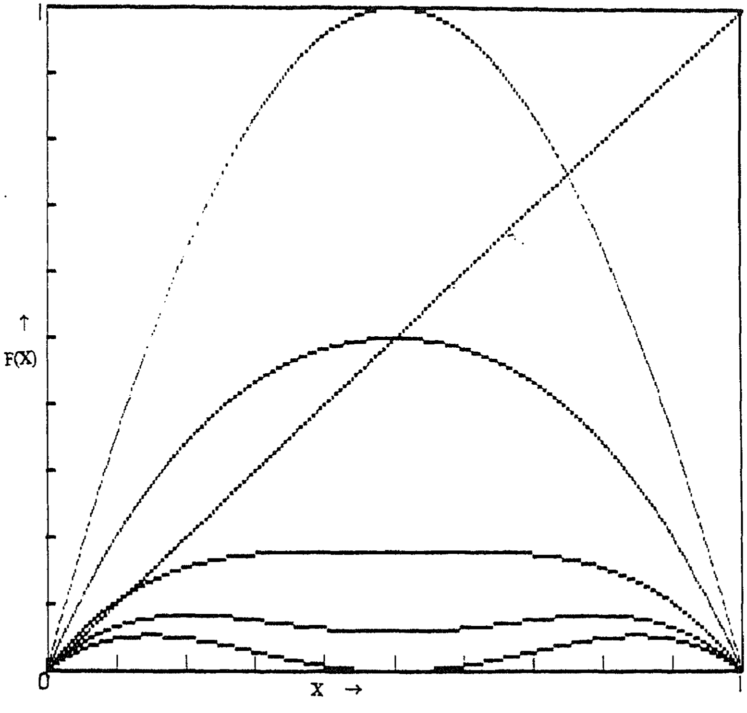


Figure 4. The combination map for $\lambda = 1$. The curves correspond to $A = 0, 0.5,\ 8/\pi^2, 3/\pi$ and 1 in that order from top to bottom.

On the right hand side both the terms belong to the same universality class, viz, the quadratic family. The motivation for studying the map (7) is to see whether the map still has all the characteristics of the quadratic family. We shall show that it has the same universal constants as the logistic map, however, the scaling index of the Lyapunov exponent at the transition point is different.

The map (7) has an extremum at $x = 1/2$ which is a second order maximum for $(\lambda - 1) < A < 8\lambda/\pi^2$ while it is a minimum for $8\lambda/\pi^2 < A < \lambda$ and there is a point of inflexion at $x = 1/2$ for $A = 8\lambda/\pi^2$. Consequently, $f(X_n, \lambda, A)$ is one humped for $(\lambda - 1) < A < 8\lambda/\pi^2$ and two humped for $8\lambda/\pi^2 < A < \lambda$. It can easily be seen that the parameter A must lie between $(\lambda - 1)$ and λ so as to keep the iterates of (7) within the unit interval (0, 1). The function $f(x, \lambda, A)$ is plotted for various values of A in figure 4. In figure 5, we give the complete bifurcation structure for the map by keeping $\lambda = 1$ and by changing A in steps of 0.001. As the parameter A is slowly tuned, the system retraces the entire period doubling route to chaos in the reverse order and finally settles down to a one-cycle for $A > 0.2435$. One can also consider negative values of A which would bring the system from a state of periodicity to a state of chaos. By changing λ, the value A_∞ of A at which the system makes transition from chaos to order and vice versa, is determined by calculating the Lyapunov exponent and by noting the value of A at which it changes sign. The thick line figure 6 represents the value of A as a function of λ. The lines parallel to it (only two are shown in figure 6) represent the bifurcation lines along which each period doubling occurs. The shaded region corresponds to the chaotic regime of the

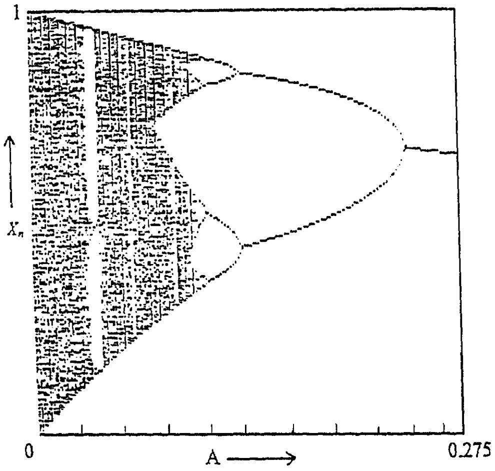


Figure 5. Bifurcation structure of the combination map with $\lambda = 1$.

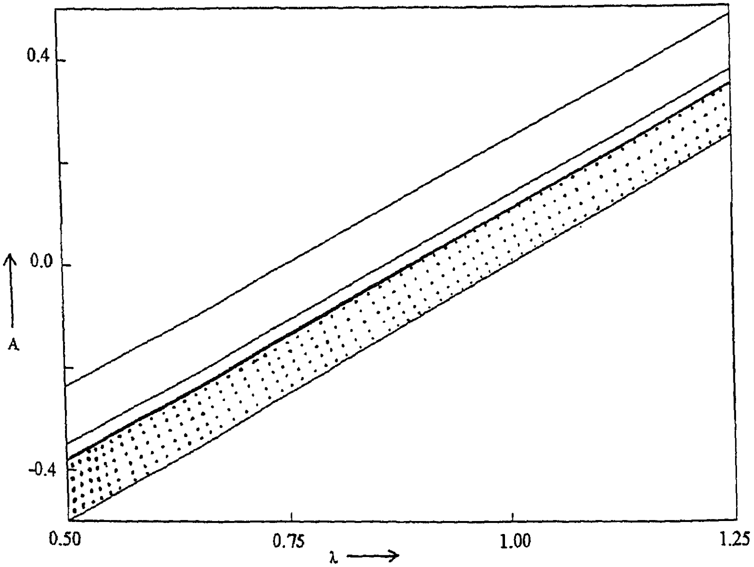


Figure 6. The parameter space (λ, A) of the combination map. The thick line represents the transition from order to chaos while the lines parallel to it are the bifurcation curves.

system where the Lyapunov exponent is generally positive except for windows of periodicity within the chaotic regime. The lowermost line represents the minimum value of A, namely $(\lambda - 1)$ for each λ. By extending this line upwards we can increase the value of λ beyond $\lambda = 1$ by taking suitably high value of A. Also, since the bifurcation curves in the (λ, A) space are parallel lines, it is clear that the Feigenbaum constant δ defined in terms of the bifurcation values A_n for fixed λ must be the same as that for the logistic map alone.

Table 2. The scaling index ν and accumulation point A_∞ for the combination map considered in § 4.

λ	A_∞	ν
0.75	−0.1382968450	0.9989838
0.875	−0.0170178750	0.9677731
1.0	−0.1046909450	0.9936844
8	7.2212675805	0.9996790

To see whether the combination map has all the characteristics of the quadratic maximum we have investigated the scaling property of the Lyapunov exponent as a function of $|A - A_\infty|$. For a one-hump map the nature of this scaling near the period doubling accumulation point has been worked out by Hubermann and Rudnick [25]. They have shown that the Lyapunov exponent σ follows the relation

$$\sigma \sim |a - a_\infty|^\nu, \tag{8}$$

where a_∞ is the accumulation point for the period doubling bifurcations and that

$$\nu = \ln 2/\ln\delta, \tag{9}$$

δ being the Feigenbaum constant. The Hubermann–Rudnick (HR) relation (9) indicates that ν depends on z, the order of the map, through the value of δ which is different for maps belonging to different universality classes (different z) [26]. Since δ is the same for the combination map one expects the same value for ν for the combination map also. To test this we have determined ν and A for the combination map and the results are shown in table 2 [24].

It is clear from the table that ν for the combination map is almost double the value of $1/2$ for the quadratic family as predicted by the HR relations and also computed numerically for the logistic map. Thus, for the combination map which also has a quadratic maximum at $x = 1/2$ does not satisfy the HR relations. Numerical studies have shown that a series of bifurcation for the chaotic band take place as $A \rightarrow A_\infty$ [23]. However, this cascade of band bifurcations does not take place ad infinitum. This incomplete nature of the cascade for the combination map could be a possible reason for the violation of HR relations for the map.

5. Concluding remarks

In this review we have considered the dynamics of modulated logistic systems. Specifically we have discussed three different systems, viz, the modulated logistic map, the parametrically perturbed logistic map and the combination map obtained by combining two maps in the same quadratic family. Each of these systems shows several interesting features, such as the enslavement of the periodicities (in MLM), occurence of superstable sequences other than the ones included in the MSS classification, the reappearance of periodicity after a chaotic regime (in the parametrically perturbed logistic map) and the violation of the Hubermann–Rudnick scaling relations for the Lyapunov exponents (in the combination map).

Acknowledgements

The author wishes to thank K P Harikrishnan, P P Saratchandran and P R Krishnan Nair for useful discussions.

References

[1] B L Hao, *Chaos – a reprint collection* (World Scientific, Singapore, 1984)
[2] J Peinke, J Parisi, O E Rossler and R Stoop, *Encounter with chaos – Self organised hierarchical complexity in semiconductor experiments* (Springer-Verlag, Berlin-Heidelberg, 1992)
[3] R M May, *Nature* (*London*), **261**, 459 (1976)
[4] M J Feigenbaum, *J. Stat. Phys.* **19**, 25 (1978)
[5] N Metropolis, M L Stein and P R Stein, *J. Comb. Theory* **A15**, 25 (1973)
[6] P T Arechi, R Meucci, G Puccioni and J Tredicce, *Phys. Rev. Lett.* **49**, 1217 (1982)
[7] R Kapral and P Mandel, *Phys. Rev.* **A32**, 1076 (1985)
[8] D Ruelle, *Proc. R. Soc.* (*London*) **A413**, 5 (1987)
[9] K P Harikrishnan and V M Nandakumaran, *Pramana – J. Phys.* **29**, 533 (1987)
[10] K P Harikrishnan and V M Nandakumaran, *Phys. Lett.* **A125**, 465 (1987)
[11] K P Harikrishnan and V M Nandakumaran, *Phys. Lett.* **A133**, 305 (1988)
[12] K P Harikrishnan and V M Nandakumaran, *Phys. Lett.* **A142**, 483 (1989)
[13] Sangeeta Batra and V S Varma, *Pramana – J. Phys.* **37**, 83 (1991)
[14] B Derrida, A Gervois and Y Pomeau, *J. Phys.* **A12**, 269 (1979)
[15] B Derrida and Y Pomeau, *Phys. Lett.* **A80**, 271 (1980)
[16] A N Sarkovskii, *Ukr. Math. J.* **16**, 61 (1964)
[17] P P Saratchandran, V M Nandakumaran and G Ambika, *Proc. of National Seminar on Computational Aspects of Chaos and Nonlinear Dynamics* edited by G Ambika and V M Nandakumaran (Wiley Eastern Ltd., 1994)
[18] S J Schiff, K Jerger, D H Duong, T Chang, M L Spano and W L Ditto, *Nature* (*London*) **370**, 615 (1994)
[19] M Lucke and Y Saito, *Phys. Lett.* **A91**, 205 (1982)
[20] Sanju and V S Varma, *Phys. Rev.* **E48** 1670 (1993)
[21] J F Heagy and S M Hammel, *Physica* **D70**, 140 (1994)
[22] T Midavainne, D Dangoisse and P Glorieux, *Phys. Rev. Lett.* **55**, 1989 (1985)
[23] P R Krishnan Nair, V M Nandakumaran and G Ambika, *Proc. of the National Seminar on Computational Aspects of Chaos and Nonlinear Dynamics* edited by G Ambika and V M Nandakumaran (Wiley Eastern Ltd., 1994)
[24] P R Krishnan Nair, V M Nandakumaran and G Ambika, *Pramana – J. Phys.* **43**, 421 (1994)
[25] B A Hubermann and J Rudnick, *Phys. Rev. Lett.* **45**, 154 (1980)
[26] G Ambika and K Babu Joseph, *Pramana – J. Phys.* **39**, 193 (1992)

Effect of boundary conditions on the invariant density of noisy maps at fully-developed chaos

V BALAKRISHNAN, G NICOLIS* and C NICOLIS†
Department of Physics, Indian Institute of Technology, Madras 600 036, India
*Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, C.P. 231, 1050 Bruxelles, Belgium
†Institut Royal Météorologique de Belgique, 1180 Bruxelles, Belgium

Abstract. The invariant density of one-dimensional maps in the regime of fully-developed chaos with uncorrelated additive noise is considered. Boundary conditions are shown to play a significant role in determining the precise form of the invariant density, via the manner in which they handle the spill-over, caused by the noise, of orbits beyond the interval. The known case of periodic boundary conditions is briefly recapitulated. Analytic solutions for the invariant density that are possible under certain conditions are presented with applications to specific well-known maps. The case of 'sticky' boundaries is generalized to 're-injection at the nearest boundary', and the exact functional equations determining the invariant density are derived. Interesting boundary layer effects are shown to occur, that lead to significant modifications of the invariant density corresponding to the unperturbed (noise-free) case, even when the latter is a constant – as illustrated by an application of the formalism to the noisy tent map. All our results are non-perturbative, and hold good for any noise amplitude in the interval.

Keywords. One-dimensional maps; fully developed chaos; invariant density; boundary conditions; noise.

PACS Nos 05.45; 05.40

1. Introduction

Random noise is almost inevitably present in systems of physical interest. Its effects on the evolution of these systems in time have long been recognized as worthy of investigation, and chaotic dynamics is no exception. From the earliest days of the modern phase in the study of deterministic chaos, there has been interest in analyzing the consequences of the inclusion of noise in chaotic flows and maps. The broad features of the resulting dynamics are understood reasonably well (see [1]–[5] and references therein).

In this paper, we re-examine the problem of one-dimensional, fully chaotic discrete dynamics with additive noise, with the aim of highlighting an aspect that has not, in our opinion, received the attention it deserves: the "spill-over" of the state variable beyond its original domain (interval) caused by the additive noise – equivalently, the effect of different boundary conditions. Studies of noisy maps have for the most part focussed on the regimes preceding fully-developed chaos, where this spill-over is perhaps not so significant, as the corresponding attractors do not comprise the entire interval. In contrast, we restrict our attention to the case of fully-developed chaos – in particular, the effect of specific boundary conditions on the invariant densities of noisy maps. In general, it is found that different boundary conditions lead to significant differences in the form of the invariant density. As the latter quantity is a central one in the statistical description of the systems concerned, we are led to the conclusion that the conjunction of noise and boundary conditions must be handled with care.

In § 2, we recapitulate the formulation of the problem along the lines best suited for what is to follow. In § 3, we consider the case of periodic boundary conditions, the one most frequently adopted in the literature [5]–[9]. After a brief summary of the formalism we point out instances where a complete solution is possible, provided certain conditions are met. In § 4, we introduce what we believe is a realistic boundary condition, namely, re-injection at the nearest boundary, derive the exact integral equation obeyed by the invariant density, and analyze the structure of its solution. A numerical illustration is provided for the case of the tent map. In § 5, we consider the generalized dimension D_q corresponding to the noisy maps investigated. Section 6 contains our concluding remarks.

2. Frobenius–Perron equation for a noisy map

Consider the one-dimensional map $x_{n+1} = f(x_n)$, $x_0 \in [a, b]$. We assume that the map is onto, i.e., we consider the case of 'fully-developed' chaos. Well-known examples include the logistic map $f(x) = 4x(1-x)$ and the tent map $f(x) = 1 - |2x-1|$ in $[0, 1]$, the square-root cusp map $f(x) = 1 - 2|x|^{\frac{1}{2}}$ in $[-1, 1]$, and so on. The invariant density $\rho^0(x)$ of the unperturbed map f is given by the Frobenius–Perron equation [10]

$$\rho^0(x) = P_0[\rho^0(x)] \equiv \int_a^b \mathrm{d}y\, \delta(x - f(y))\rho^0(y). \tag{2.1}$$

We assume that the operator P_0 is known explicitly (via the functional equation to which (2.1) is reducible), as also the solution $\rho^0(x)$.

Now consider the noisy map

$$x_{n+1} = f(x_n) + \xi_n, \tag{2.2}$$

where $\langle \xi_n \rangle = 0$, $\langle \xi_n \xi_{n'} \rangle = \langle \xi^2 \rangle \delta_{nn'}$ (uncorrelated noise). We want to find an exact expression for the invariant density in this case, given the distribution $g(\xi)$ of the noise. Earlier treatments either ignore the complications due to the 'spill-over' of $f(x_n) + \xi_n$ in some realizations to values outside the interval $[a, b]$, or else assume [5]–[9] that $g(\xi)$ is a periodic function with fundamental interval $[a, b]$. ($g(\xi)$ is generally taken to be a Gaussian with zero mean, or an infinite sum of Gaussians to ensure periodicity.) Let us denote by P the perturbed Frobenius–Perron operator (i.e., the one corresponding to the noisy map), and by $\rho(x)$ the corresponding invariant density, averaged over the realizations of the additive noise. A formal expression for the kernel of the operator P is well known, and is obtained as follows. From (2.2), it follows that the noise-averaged density at time n evolves according to

$$\rho_{n+1} = \int \mathrm{d}\xi \int_a^b \mathrm{d}y\ \delta(x - f(y) - \xi)\ \ \rho_n(y)\ g(\xi). \tag{2.3}$$

This is an exact expression if the noise distribution is periodic, but imprecise if it is not, precisely because the spill-over across the boundaries a and b has been ignored (nor is it immediately clear what the effects of this neglect will be). Formally performing the integration over ξ using the δ-function, one gets for the invariant density the perturbed Frobenius–Perron [3, 6, 10]

$$\rho(x) = P[\rho(x)] \equiv \int_a^b \mathrm{d}y\, g(x - f(y))\, \rho(y). \tag{2.4}$$

Comparision with the noise-free case, (2.1), shows that the effect of the noise is to replace the δ-function kernel of P_0 with the noise distribution function g.

3. Periodic boundary conditions

3.1 *Fourier expansion method*

The convolution form of (2.4) suggests a solution in terms of a Fourier expansion, and this is what has been done in earlier treatments [6, 7], taking g to be a periodic function of its argument. As mentioned earlier, this is appropriate when periodic boundary conditions are assumed. We may then write

$$g(\xi) = (b-a)^{-1} \sum_{n=-\infty}^{\infty} \tilde{g}_n \exp(ik_n\xi), \tag{3.1}$$

where $k_n = 2\pi n/(b-a)$. The inversion formula is

$$\tilde{g}_n = \int_a^b \mathrm{d}\xi\, g(\xi) \exp(-ik_n\xi). \tag{3.2}$$

Expanding $\rho(x)$ in a similar fashion, i.e.,

$$\rho(\xi) = (b-a)^{-1} \sum_{n=-\infty}^{\infty} \tilde{\rho}_n \exp(ik_n\xi), \tag{3.3}$$

we obtain from (2.4) the relation

$$\tilde{\rho}_n = \sum_{m=-\infty}^{\infty} \tilde{g}_n S_{nm} \tilde{\rho}_m, \tag{3.4}$$

where

$$S_{nm} = (b-a)^{-1} \int_a^b \mathrm{d}y \exp i[k_m y - k_n f(y)]. \tag{3.5}$$

The effects of the noise are contained in $\{\tilde{g}_n\}$. The noise-free case is recovered by setting $\tilde{g}_n = 1$ for all n, corresponding to replacing $g(\xi)$ by $\delta(\xi)$. The normalization conditions on g and ρ imply that

$$\int_a^b g(\xi)\mathrm{d}\xi = \tilde{g}_0 = 1, \quad \int_a^b \rho(x)\mathrm{d}x = \tilde{\rho}_0 = 1. \tag{3.6}$$

Equation (3.4) can be written [6, 7] as the following *inhomogeneous* equation for $\tilde{\rho}_n$, $n \neq 0$:

$$\tilde{\rho}_n = \tilde{g}_n S_{n0} + \sum_{m=-\infty}^{\infty}{}' \tilde{g}_n S_{nm} \tilde{\rho}_m, \quad (n \neq 0) \tag{3.7}$$

where $\sum'$ is over non-zero integers m. In principle, iteration of (3.7) yields the formal solution

$$\tilde{\rho}_n = \tilde{g}_n S_{n0} + \sum_m{}' \tilde{g}_n \tilde{g}_m S_{nm} S_{m0} + \sum_m{}' \sum_l{}' \tilde{g}_n \tilde{g}_m \tilde{g}_l S_{nm} S_{ml} S_{l0} + \cdots \tag{3.8}$$

for $\tilde{\rho}_n$, and hence for $\rho(x)$. Rapid convergence of the iteration, at least in the fully chaotic cases considered here, helps numerical calculations in specific instances [6, 7]. Clearly, it is rather difficult to extract analytical information regarding the dependence of $\rho(x)$ on the parameters of the noise distribution from the formal solution, except under special circumstances. To reconstruct $\rho(x)$ in closed form, one must find $\tilde{\rho}_n$ exactly and then sum the Fourier series. In fact, this procedure is complicated enough even in the noise-free case ($\tilde{g}_n = 1$) – e.g., for the logistic map $x_{n+1} = 4x_n(1 - x_n)$ on $[0, 1]$, S_{nm} is expressible [6] in terms of Fresnel integrals, but finding $\tilde{\rho}_n$ explicitly and summing the Fourier series hardly seems to be the easiest way to obtain the well-known result [11] $\rho^0(x) = [\pi^2 x(1 - x)]^{-\frac{1}{2}}$. However, there are two situations in which (3.4) (or (3.7)) can be solved easily. These are considered below.

3.2 $S_{n0} = 0$; *application to the Bernoulli and tent maps*

If the quantity

$$S_{n0} = (b - a)^{-1} \int_a^b \mathrm{d}y \exp\left[-ik_n f(y)\right] \tag{3.9}$$

happens to vanish for every nonzero integer n, then (3.7) reduces to a set of *homogeneous* equations for $\{\tilde{\rho}_n\}$, $n \neq 0$. The determinant of this set of equations will be nonzero in general, so that only the trivial solution $\tilde{\rho}_n = 0$, $n \neq 0$, exists. This means that

$$\rho(x) = \rho^0(x) = (b - a)^{-1} \tag{3.10}$$

in this case, for any (periodic) noise distribution $g(\xi)$. This is what happens for the general Bernoulli map

$$x_{n+1} = \begin{cases} \alpha x_n\,, & 0 \leq x_n \leq 1/\alpha \\ (\alpha x_n - 1)/(\alpha - 1)\,, & 1/\alpha < x_n \leq 1\,, \end{cases} \tag{3.11}$$

where $\alpha > 1$. The same result holds good for the continuous counterpart of (3.11), namely, the tent map in which the second segment of the map reads $x_{n+1} = \alpha(1 - x_n)/(\alpha - 1)$. (For $\alpha = 2$, these maps are the usual Bernoulli map and the symmetric tent map, respectively). Another example, involving a map that is *not* piecewise linear, is the antisymmetric square-root cusp map in $[-1, 1]$ given by

$$x_{n+1} = \begin{cases} 1 - 2\sqrt{-x_n}\,, & -1 \leq x_n \leq 0 \\ 2\sqrt{x_n} - 1, & 0 < x_n \leq 1\,. \end{cases} \tag{3.12}$$

We find in this case

$$S_{nm} = (-1)^n \int_0^1 dy \cos\, \pi\,(my - 2n\sqrt{y}), \tag{3.13}$$

so that $S_{n0} = 0$. This leads, as before, to the constant normalized density $\rho(x) = \rho^0(x) = 1/2$ for both the unperturbed map and the noisy map. In general: under periodic boundary conditions, additive white noise does not alter the unperturbed invariant density ρ^0 when the latter is a constant. We shall see later that this conclusion does *not* hold good if the boundary conditions are changed.

3.3 $S_{nm} = \pm S_{n,-m}$, $S_{n0} = \mp S_{-n,0}$: *application to the cusp map*

There is another situation in which an explicit solution for $\rho(x)$ is feasible in the above approach, even if $\rho^0(x)$ is not a constant (i.e., $S_{n0} \neq 0$). This occurs when (i) S_{nm} is even [odd] in the index m, (ii) S_{n0} is odd [even] in the index n, and further (iii) $\tilde{g}_n = \tilde{g}_{-n}$, i.e., $\tilde{g}_n$ is real (which happens when $g(\xi)$ is symmetric about $(b-a)/2$). It is evident that if conditions (i)–(iii) are satisfied, all terms beyond the first on the right-hand side of the iterative solution (3.8) vanish identically. Moreover, $\tilde{\rho}_n^0 = S_{n0}$ in this case (since the noise-free case corresponds to $\tilde{g}_n = 1$). We are therefore enabled to arrive at the solution

$$\tilde{\rho}_n = \tilde{g}_n S_{n0} = \tilde{g}_n \tilde{\rho}_n^0. \tag{3.14}$$

Therefore

$$\rho(x) = (b-a)^{-1} \sum_{n=-\infty}^{\infty} \tilde{g}_n S_{n0} \exp(ik_n x). \tag{3.15}$$

Inserting the definitions of $\tilde{g}_n$ and S_{n0}, and using the relation

$$(b-a)^{-1} \sum_{n=-\infty}^{\infty} \exp(ik_n z) = \delta(z), \tag{3.16}$$

we find the solution

$$\rho(x) = (b-a)^{-1} \int_a^b d\xi \int_a^b dy\; g(\xi)\; \delta(x - f(y) - \xi). \tag{3.17}$$

A concrete example of the situation just described is provided by the symmetric square-root cusp map

$$x_{n+1} = 1 - 2|x_n|^{\frac{1}{2}}, \quad x \in [-1, 1]. \tag{3.18}$$

This map is of importance as a model of intermittent chaos, arising from the marginal stability of the fixed point at the left boundary $x = -1$. This is reflected in its invariant density [12]

$$\rho^0(x) = (1-x)/2, \tag{3.19}$$

which shows how the probability is 'piled up' towards $x = -1$. We have in this case

$$S_{nm} = (-1)^n \int_0^1 dy \cos(\pi m y) \exp(2\pi n i \sqrt{y}), \tag{3.20}$$

counterpart (3.12) of the cusp map). Thus S_{nm} is even in m. Further,

$$S_{n0} = (-1)^n (i\pi n)^{-1}, \tag{3.21}$$

which is odd in n. Therefore the conditions just discussed are met, and we have $\tilde{\rho}_n^0 = S_{n0}$, for $n \neq 0$. Together with $\tilde{\rho}_0^0 = 1$, this leads on summing the Fourier series to the recovery of the known result (3.19) for the invariant density $\rho^0(x)$ of the unperturbed map. The point we now make is that this form of the invariant density can undergo significant modification in the presence of noise. For a symmetric noise distribution in $[-1,1]$ (so that $\tilde{g}_n = \tilde{g}_{-n}$), using (3.6), (3.14) and (3.21) we get the exact result

$$\rho(x) = \frac{1}{2} + \sum_{n=1}^{\infty} (-1)^n \, \tilde{g}_n \, \frac{\sin n\pi x}{n\pi}. \tag{3.22}$$

The 'antisymmetry' of $\rho(x)$ about the value $\frac{1}{2}$ is at once evident, as is the fact that $\rho(0) = \frac{1}{2}$. Further, if $\rho(-1) = \rho(1)$, then each of these is also equal to $\frac{1}{2}$; else, their mean value is equal to $\frac{1}{2}$. In order to exhibit analytically the form of the solution, let us consider first the simple case of a rectangular or uniform noise, given by the density

$$g(\xi) = \frac{1}{2\eta} \theta(\eta - |\xi|) \tag{3.23}$$

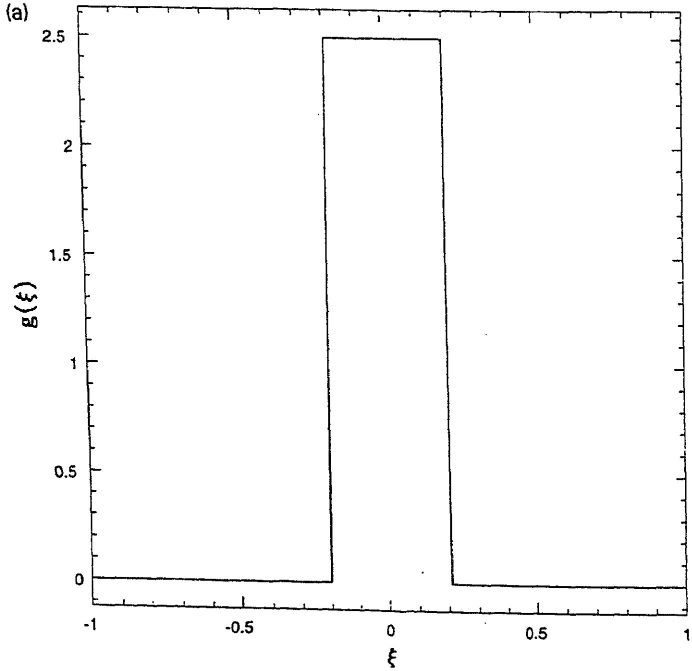


Figure 1(a). The rectangular (uniform) noise distribution $g(\xi) = (2\eta)^{-1}\theta(\eta - |\xi|)$, for $\eta = 0.2$.

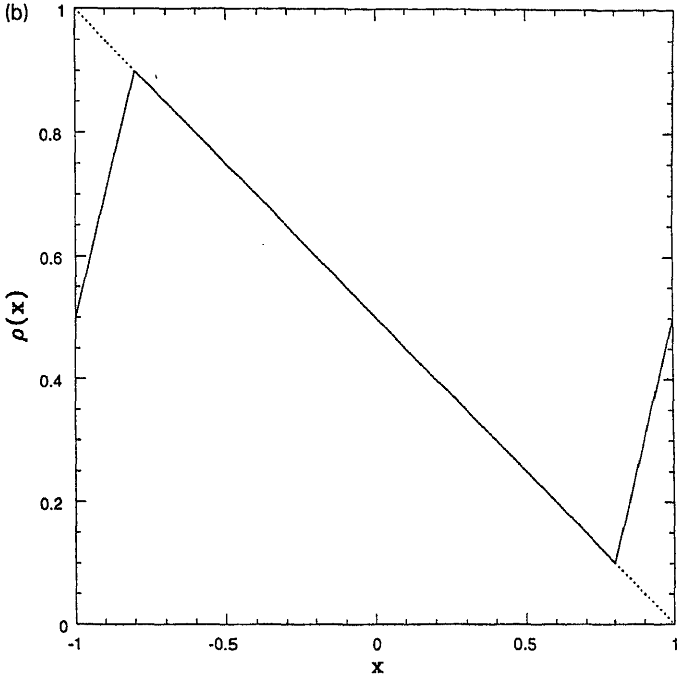


Figure 1(b). The invariant density $\rho(x)$ for the symmetric cusp map $f(x) = 1 - 2|x|^{\frac{1}{2}}$ in the presence of additive noise distributed as in figure 1(a), under periodic boundary conditions. The dotted line depicts the invariant density $(1-x)/2$ of the unperturbed (noise-free) map.

in the fundamental interval $[-1, 1]$, where $0 < \eta < 1$. The amplitude of the noise is thus limited by the parameter η. Carrying out the calculations required, we arrive at the result

$$\rho(x) = \begin{cases} \frac{1}{2}[\eta^{-1} + (\eta^{-1} - 1)x]\,, & -1 \le x \le -1 + \eta \\ \frac{1}{2}[1 - x]\,, & -1 + \eta \le x \le 1 - \eta \\ \frac{1}{2}[(2 - \eta^{-1}) + (\eta^{-1} - 1)x]\,, & 1 - \eta \le x \le 1\,. \end{cases} \qquad (3.24)$$

The noise distribution (3.23) and the invariant density are shown in figures 1(a) and (b) respectively. We note how the noise suppresses the effect of the marginally stable fixed point of the map at $x = -1$. Equation (3.24) is an exact result (under the periodic boundary conditions assumed) that is valid, in fact, for any value of the amplitude η of the noise. As $\eta \to 1$, $\rho(x)$ tends to the uniform distribution $\rho(x) = 1/2$, as one might expect under the circumstances.

The fact that $\rho(x)$ (3.24) continues to remain piecewise linear is actually an artefact of the uniform noise distribution (3.23). A smoother modification of the unperturbed density $(1-x)/2$, but one that retains the general features of $\rho(x)$ as found above, is provided by the one-parameter family of unimodal, normalized noise densities

$$g(\xi;\mu) = \frac{\mu(\cos\mu\xi - \cos\mu)}{2(\sin\mu - \mu\cos\mu)}\,, \quad \xi \in [-1, 1], \qquad (3.25)$$

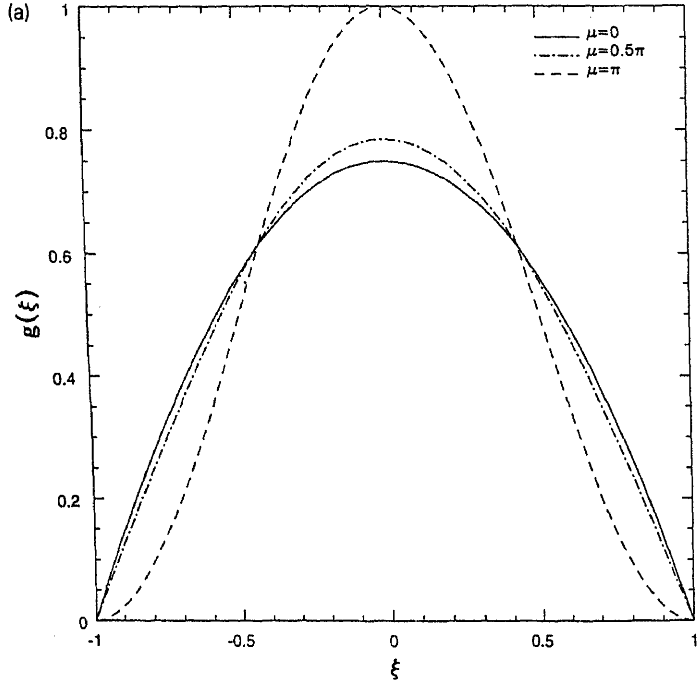


Figure 2(a). The noise distribution $g(\xi;\mu)$ of (3.25) in the limiting cases $\mu=\pi$ (3.27) and $\mu=0$ (3.29). Also shown is the case $\mu=\pi/2$, i.e., $g(\xi;\pi/2)=(\pi/4)\cos(\pi\xi/2)$.

where $0<\mu\leq\pi$. We then find the corresponding invariant density to be given by

$$\rho(x)=\frac{1}{2}\left[1+(sgn\ x)\frac{\sin\mu(1-|x|)-(1-|x|)\sin\mu}{(\mu\cos\mu-\sin\mu)}\right]. \tag{3.26}$$

Thus $\rho(-1)=\rho(0)=\rho(1)=1/2$. In the limiting case $\mu=\pi$, we have

$$g(\xi;\pi)=\tfrac{1}{2}(1+\cos\pi\xi), \tag{3.27}$$

the corresponding invariant density being

$$\rho(x)=\frac{1}{2}\left(1-\frac{\sin\pi x}{\pi}\right). \tag{3.28}$$

The limit $\mu=0$ is exceptional. It corresponds to the parabolic distribution

$$g(\xi;0)=\tfrac{3}{4}(1-\xi^2), \tag{3.29}$$

leading to the invariant density

$$\rho(x)=\tfrac{1}{2}(1-x)+\tfrac{3}{4}x(|x|-x^2). \tag{3.30}$$

In this case, $\rho(-1)$ and $\rho(1)$ remain equal to 1 and 0 respectively. The noise distributions of (3.27) and (3.29) are shown in figure 2(a), while the corresponding solutions for $\rho(x)$

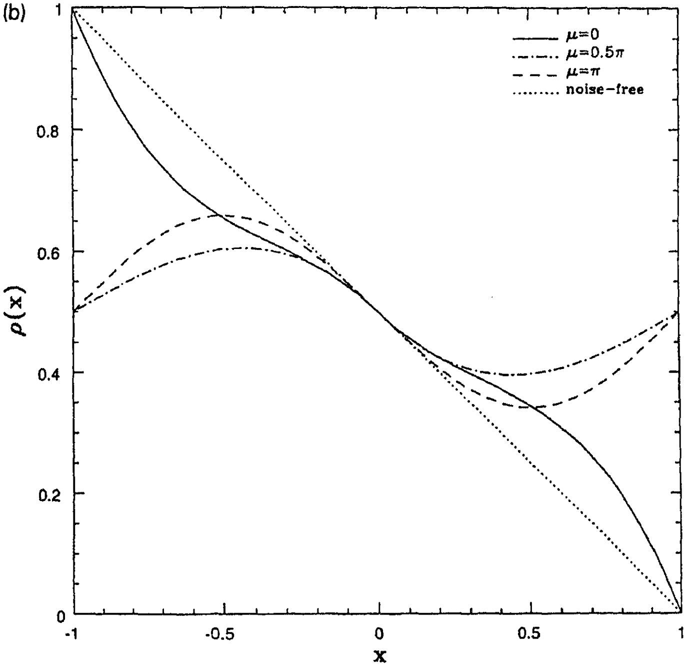


Figure 2(b). The corresponding invariant densities (3.26) for the noisy, symmetric cusp map, given by (3.28) for $\mu = \pi$ and by (3.30) for $\mu = 0$. Also shown is the case $\mu = \pi/2$, for which $\rho(x) = \frac{1}{2}(2 - x - \cos(\pi x/2))$ $(x > 0)$ and $\rho(x) = \frac{1}{2}(\cos(\pi x/2) - x)$ $(x < 0)$.

are depicted in figure 2(b). The 'evening out' effect of the noise is quite evident. It must be borne in mind, though, that these results correspond to relatively large stochastic perturbations of the original dynamics: the standard deviation of the noise for the distribution (3.29) is $1/\sqrt{5} \approx 0.45$, while that for (3.27) is $\sqrt{(1/3) - (2/\pi^2)} \approx 0.36$. The Lyapunov exponent λ, which is equal to $1/2$ for the unperturbed map [12], is also altered in the presence of noise: we have $\ln |f'(x)| = -\frac{1}{2} \ln |x|$, so that the second term on the right in eq. (3.22) does not contribute to λ, being odd in x.

3.4 *Functional equation under uniformly distributed noise*

We present in brief a formalism that avoids Fourier series and leads to a functional equation for the invariant density, for an onto map in the fundamental interval $[a, b]$ perturbed by additive noise ξ, under periodic boundary conditions. This may be a useful approach in numerical calculations, especially in situations in which the noise is not just a small perturbation.

Returning to (2.3), we insert the factor $1 = \int_a^b \mathrm{d}u\, \delta(u - f(y))$ on the right-hand side:

$$\rho_{n+1}(x) = \int \mathrm{d}\xi \int_a^b \mathrm{d}y \int_a^b \mathrm{d}u\, \delta(x - f(y) - \xi)\delta(u - f(y))\, g(\xi)\, \rho_n(y)$$

$$= \int \mathrm{d}\xi \int_a^b \mathrm{d}y\, \delta(x-u-\xi) g(\xi) \int_a^b \mathrm{d}u\, \delta(u-f(y))\ \rho_n(y)$$

$$= \int \mathrm{d}\xi \int_a^b \mathrm{d}u\, \delta(x-u-\xi)\ g(\xi)\ P_0[\rho_n(u)], \tag{3.31}$$

using the definition of the unperturbed Frobenius–Perron operator P_0. Therefore the invariant density is given by

$$\rho(x) = \int \mathrm{d}\xi \int_a^b \mathrm{d}u\, \delta(x-u-\xi)\ g(\xi)\ P_0[\rho(u)]. \tag{3.32}$$

For brevity, let us introduce the notation $P_0[\rho(u)] = \bar{\rho}(u)$. (Of course, $\bar{\rho}^0(x) = \rho^0(x)$ itself). Thus

$$\rho(x) = \int \mathrm{d}\xi \int_a^b \mathrm{d}u\, \delta(x-u-\xi)\ g(\xi)\bar{\rho}(u). \tag{3.33}$$

For the sake of definiteness (so as to be able to write all formulas explicitly), we consider henceforth the case of uniformly distributed noise, as in (3.23), with $0 < \eta < (b-a)/2$. For periodic boundary conditions, the support of the δ-function in (2.3) must be $(f(y)+\xi) \bmod (b-a)$, i.e., the support of $\delta(x-u-\xi)$ in (3.33) is $u = (x-\xi) \bmod (b-a)$. Writing $\delta^{\bmod}(x-u-\xi)$ to keep track of this fact, we have

$$\rho(x) = (2\eta)^{-1} \int \mathrm{d}\xi\, \theta(\eta - |\xi|) \int_a^b \mathrm{d}u\, \delta^{\bmod}(x-u-\xi)\bar{\rho}(u). \tag{3.34}$$

Carrying out the integration over ξ, we find:

$$\rho(x) = \begin{cases} (2\eta)^{-1}\left[\displaystyle\int_a^{x+\eta} \mathrm{d}u\, \bar{\rho}(u) + \int_{x-\eta+b-a}^{b} \mathrm{d}u\, \bar{\rho}(u)\right], & \text{for } a \le x < a+\eta, \\ (2\eta)^{-1}\left[\displaystyle\int_{x-\eta}^{x+\eta} \mathrm{d}u\ \bar{\rho}(u)\right], & \text{for } a+\eta \le x \le b-\eta, \\ (2\eta)^{-1}\left[\displaystyle\int_{x-\eta}^{b} \mathrm{d}u\, \bar{\rho}(u) + \int_a^{x+\eta-(b-a)} \mathrm{d}u\, \bar{\rho}(u)\right], & \text{for } b-\eta < x \le b. \end{cases} \tag{3.35}$$

Equation (3.35) must be solved numerically to find $\rho(x)$.

It is evident that $\rho(x)$ is continuous at all points in the interval, including the points $a+\eta$ and $b-\eta$. A convenient way of writing the functional equation (3.35) is in terms of the derivative of $\rho(x)$: we have

$$2\eta\rho'(x) = \begin{cases} \bar{\rho}(x+\eta) - \bar{\rho}(x-\eta+b-a), & a \le x < a+\eta, \\ \bar{\rho}(x+\eta) - \bar{\rho}(x-\eta), & a+\eta \le x \le b-\eta, \\ \bar{\rho}(x-\eta) - \bar{\rho}(x+\eta-b+a), & b-\eta < x \le b. \end{cases} \tag{3.36}$$

The discontinuity of $\rho'(x)$ at $x = a+\eta$ and $x = b-\eta$ is given by

$$\mathrm{disc}\ \rho'(a+\eta) = -\mathrm{disc}\ \rho'(b-\eta) = (2\eta)^{-1}[\bar{\rho}(b) - \bar{\rho}(a)]. \tag{3.37}$$

This relation may be verified for the cusp map using the solution (3.24) already found for the invariant density.

4. Sticky boundaries: Re-injection at the nearest boundary

4.1 *General equation for the invariant density*

While periodic boundary conditions constitute a convenient way of handling the spill-over problem in a noisy map, other solutions are also possible [5]. These include reflecting boundary conditions, or working in the infinite domain $(-\infty,\infty)$ so that natural boundary conditions can be employed. Yet another approach is to tailor the noise distribution by making it state-dependent, such that the probability of an escape of the dynamical variable out of the interval is zero. This means that the noise is actually multiplicative rather than additive, with a density $g(\xi,x)$ satisfying $g(\xi,a)=g(\xi,b)=0$. One way of implementing this is to assume $g(\xi,x)$ to be of the form [13]

$$g(\xi,x)=\begin{cases} g(\xi)/\int_a^b g(\xi'-x)\mathrm{d}\xi', & x+\xi\in[a,b] \\ 0, & \text{otherwise}. \end{cases} \tag{4.1}$$

While it is conceivable that this and other such prescriptions might be justified on physical grounds in certain instances, the very modelling of the stochastic forces on a system by means of multiplicative noise may be quite inappropriate in others, particularly if the noise arises from sources that are quite decoupled from the system dynamics.

A model that appears to be rather more plausible on physical grounds is provided by 'sticky' boundaries: once the dynamical system reaches a boundary value, it stays put at that value, until the noise moves it away. More generally, instead of escape out of the interval owing to the addition of noise to the evolution, we may assume re-injection at the nearest boundary. This avoids the arbitrary clipping of the orbit of x_n when $f(x_n)+\xi$ goes out of $[a,b]$ by re-defining the noisy map as follows. Denoting $f(x_n)+\xi$ by $F(x_n,\xi_n)$, we have

$$x_{n+1}=\begin{cases} f(x_n)+\xi_n & \text{if } a\le F(x_n,\xi_n)\le b, \\ a & \text{if } F(x_n,\xi_n)\le a, \\ b & \text{if } F(x_n,\xi_n)\ge b. \end{cases} \tag{4.2}$$

Therefore if the representative point overshoots the boundary at a or at b, it is replaced at that boundary, remaining trapped there until it is re-injected into the phase space by the noisy map. (This is so even if a or b is a fixed point of the unperturbed map.) The end points thus act as temporary traps that prevent any loss of 'measure' due to overflow.

As before, we are interested in the invariant density $\rho(x)$, rather than the time-dependent density $\rho_n(x)$. The counterpart of (2.3) for the map (4.2) is, in the limit $n\longrightarrow\infty$,

$$\rho(x)=\int \mathrm{d}\xi\int_a^b [\delta(x-F(y,\xi)) \\ +\delta(x-a)\theta(a-F(y,\xi))+\delta(x-b)\theta(F(y,\xi)-b)]\rho(y)g(\xi). \tag{4.3}$$

We note that the re-injection prescription in (4.2) leads, in general, to δ-functions in the density with support at the boundary points. Hence $\rho(x)$ is of the form

$$\rho(x) = \alpha\delta(x-a) + \beta\delta(x-b) + \rho_*(x), \tag{4.4}$$

where α and β are (dimensionless) constants, and $\rho_*(x)$ is the 'regular' part of the invariant density. Substituting (4.4) in (4.3), we get the coupled equations

$$\alpha = \int \mathrm{d}\xi\, g(\xi)\Big[\alpha\theta(a - F(a,\xi)) + \beta\theta(a - F(b,\xi)) + \int_a^b \mathrm{d}y\, \theta(a - F(y,\xi))\rho_*(y)\Big], \tag{4.5}$$

$$\beta = \int \mathrm{d}\xi\, g(\xi)\Big[\alpha\theta(F(a,\xi) - b) + \beta\theta(F(b,\xi) - b) + \int_a^b \mathrm{d}y\, \theta(F(y,\xi) - b)\rho_*(y)\Big], \tag{4.6}$$

$$\rho_*(x) = \int \mathrm{d}\xi\, g(\xi)\Big[\alpha\delta(x - F(a,\xi)) + \beta\delta(x - F(b,\xi)) + \int_a^b \mathrm{d}y\, \delta(x - F(y,\xi))\rho_*(y)\Big]. \tag{4.7}$$

As before, in order to be specific, we shall take $g(\xi)$ to be the amplitude-limited (uniform) noise distribution of (3.23). Further, let us assume that the map f is a continuous, unimodal, onto map, with $f(y)$ increasing monotonically from $y = a$ to $y = y_m$, and decreasing monotonically from y_m to b; $f(a) = f(b) = a, f(y_m) = b$. (The modifications required in other cases can be worked out.) Let $y_{1,2}(z)$ denote the two roots of $f(y) = z \in [a,b]$. We have $y_1(a) = a$, $y_2(a) = b$, $y_{1,2}(b) = y_m$. Then, after some simplification (the details are described in the Appendix), we arrive at the following equations for α, β and $\rho_*(x)$:

$$\alpha = \beta + \frac{1}{\eta}\left\{\int_a^{y_1(a+\eta)} \mathrm{d}y(a+\eta-f(y))\rho_*(y) + \int_{y_2(a+\eta)}^b \mathrm{d}y(a+\eta-f(y))\rho_*(y)\right\}, \tag{4.8}$$

$$\beta = \frac{1}{2\eta}\int_{y_1(b-\eta)}^{y_2(b-\eta)} \mathrm{d}y(\eta - b + f(y))\ \rho_*(y), \tag{4.9}$$

$$\rho_*(x) = \left(\frac{\alpha+\beta}{2\eta}\right)[\theta(x-a) - \theta(x-a-\eta)] + \frac{1}{2\eta}\int_{-\eta}^{\eta}\mathrm{d}\xi \int_a^b \mathrm{d}y\, \delta(x - f(y) - \xi)\rho_*(y). \tag{4.10}$$

In the last integral on the right, the factor $1 = \int_a^b \mathrm{d}u\ \delta(u - f(y))$ may be inserted as before, to lead to

$$\rho_*(x) = \left(\frac{\alpha+\beta}{2\eta}\right)[\theta(x-a) - \theta(x-a-\eta)] + \frac{1}{2\eta}\int_{\max(a,x-\eta)}^{\min(b,x+\eta)} \mathrm{d}u\, \bar{\rho}_*(u). \tag{4.11}$$

$$\alpha+\beta+\int_a^b \rho_*(x)\mathrm{d}x = 1. \tag{4.12}$$

We note that $\alpha > \beta$. Moreover, in addition to the δ-function spikes at the boundaries, there occurs an extra boundary layer contribution on the left, in $(a, a+\eta)$. There is a discontinuity in $\rho(x)$ at $x = a+\eta$, and a discontinuous change in slope at $x = b-\eta$, with the respective discontinuities given by

$$\mathrm{disc}\,\rho(a+\eta) = -\left(\frac{\alpha+\beta}{2\eta}\right), \tag{4.13}$$

$$\mathrm{disc}\,\rho'(b-\eta) = -\bar{\rho}_*(b)/(2\eta) = -\frac{1}{4\eta^2}\int_{b-\eta}^{b} \mathrm{d}u\;\bar{\rho}_*(u). \tag{4.14}$$

The Lyapunov exponent is also altered, in general:

$$\lambda = \alpha\ln|f'(a)| + \beta\ln|f'(b)| + \int_a^b \rho_*(x)\ln|f'(x)|\,\mathrm{d}x. \tag{4.15}$$

We illustrate the foregoing in the case of the tent map at fully-developed chaos.

4.2 *Application to the tent map*

The map concerned is $f(x) = 1-|2x-1|$, $x \in [0,1]$. Thus $a = 0$, $b = 1$, $y_m = \frac{1}{2}$, $y_1(z) = z/2$, $y_2(z) = 1-(z/2)$. Further,

$$P_0[\rho(u)] \equiv \bar{\rho}(u) = \frac{1}{2}\left[\rho\left(\frac{u}{2}\right) + \rho\left(1-\frac{u}{2}\right)\right],$$

and the noise-free map has the constant invariant density $\rho^0(x) = 1$. The noisy map has an invariant density of the form

$$\rho(x) = \alpha\delta(x) + \beta\delta(x-1) + \rho_*(x), \tag{4.16}$$

where

$$\alpha = \beta + \frac{1}{\eta}\left\{\int_0^{\eta/2} \mathrm{d}y(\eta-2y)\rho_*(y) + \int_{(1-\eta/2)}^{1} \mathrm{d}y\,(\eta-2+2y)\rho_*(y)\right\}, \tag{4.17}$$

$$\beta = \frac{1}{2\eta}\int_{(1-\eta)/2}^{(1+\eta)/2} \mathrm{d}y\,(\eta-|2y-1|)\;\rho_*(y), \tag{4.18}$$

and

$$\rho_*(x) = \left(\frac{\alpha+\beta}{2\eta}\right)[\theta(x)-\theta(x-\eta)] + \frac{1}{2\eta}\int_{\max(0,x-\eta)}^{\min(1,x+\eta)} \mathrm{d}y\,\bar{\rho}_*(y). \tag{4.19}$$

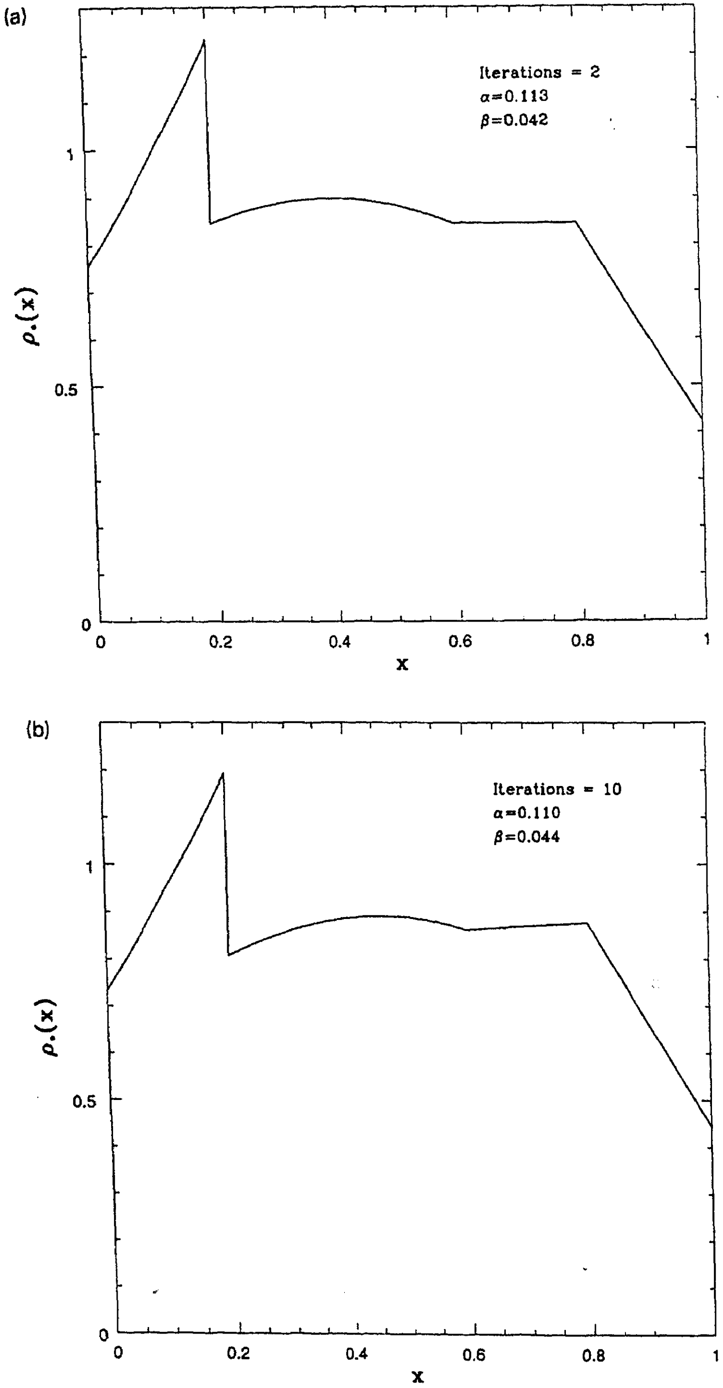


Figure 3(a–b). *(Continued)*

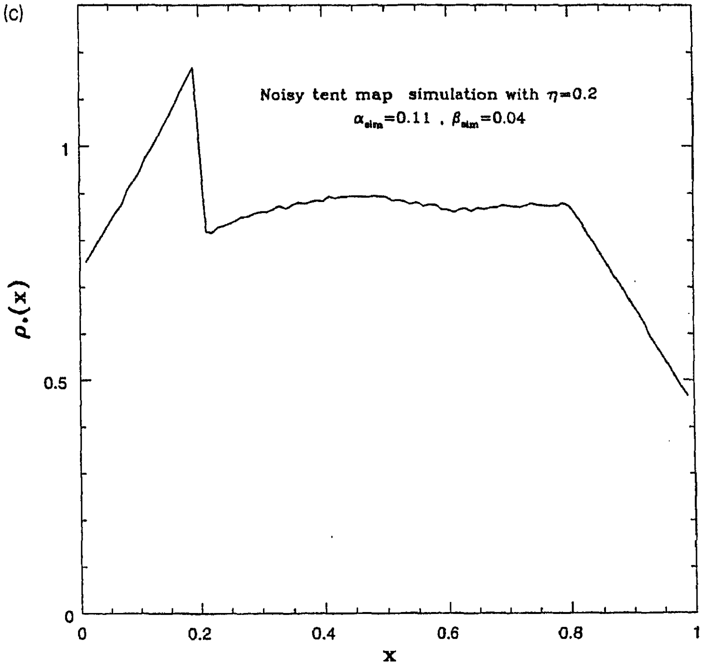


Figure 3. Numerical solution and comparision with simulation for the invariant density $\rho(x) = \alpha\delta(x) + \beta\delta(1-x) + \rho_*(x)$ of the noisy tent map with noise amplitude $\eta = 0.2$. (**a**) Solution after 2 iterations, starting with the zeroth approximation $\rho_*(x) = 1$. The solution essentially stablizes (correct to 3 decimal places) in 3 iterations. (**b**) Solution after 10 iterations. (**c**) The result of a numerical simulation on the noisy tent map with the re-injection prescription, (4.2). The result shown is an average over 10 simulations of 10^6 iterations each, with a bin size of 10^{-2}.

Equation (4.19) yields the following, for x in different ranges:

$$0 \le x < \eta: \;\; \rho_*(x) = \frac{\alpha+\beta}{2\eta} + \frac{1}{2\eta}\left[\int_0^{(x+\eta)/2} \mathrm{d}y\rho_*(y) + \int_{1-(x+\eta)/2}^{1} \mathrm{d}y\,\rho_*(y)\right], \tag{4.20}$$

$$\eta \le x < 1-\eta: \;\; \rho_*(x) = \frac{1}{2\eta}\left[\int_{(x-\eta)/2}^{(x+\eta)/2} \mathrm{d}y\rho_*(y) + \int_{1-(x+\eta)/2}^{1-(x-\eta)/2} \mathrm{d}y\rho_*(y)\right], \tag{4.21}$$

$$1-\eta \le x \le 1: \;\; \rho_*(x) = \frac{1}{2\eta}\int_{(x-\eta)/2}^{1-(x-\eta)/2} \mathrm{d}y\,\rho_*(y). \tag{4.22}$$

We note right away that $\rho(x) = 1$ or even $\rho_*(x) =$ constant is not a solution, although $\rho^0(x) = 1$. Equations (4.17)–(4.22) must be solved numerically. The broad features of the solution can be seen by adopting an iterative procedure: if we take $\rho_*(x) = 1$ in the zeroth approximation, we obtain the following as the first approximation (after normalization):

$$\alpha = \frac{3\eta}{4(1+\eta)}, \quad \beta = \frac{\eta}{4(1+\eta)}, \tag{4.23}$$

$$\rho_*(x) = \begin{cases} (1+\frac{x}{2\eta})/(1+\eta), & 0 \le x < \eta \\ 1/(1+\eta), & \eta \le x \le 1-\eta \\ (\frac{1}{2}+\frac{1-x}{2\eta})/(1+\eta), & 1-\eta \le x \le 1. \end{cases} \tag{4.24}$$

Successive iterations push the 'plateau' in $\rho_*(x)$ further to the right. The actual solution does not have any interval in which $\rho_*(x)$ is a constant, but rather a range $(\eta, 1-\eta)$ in which its variation with x is gentle. It varies linearly in the ranges $(0,\eta)$, $((1+\eta)/2, 1-\eta)$ and $(1-\eta, \eta)$. Figures 3(a) and 3(b) show how the solution for $\rho_*(x)$ stabilizes as the set (4.17)–(4.22) is solved iteratively, in the case $\eta = 0.2$. The discontinuity in $\rho_*(x)$ at $x = \eta = 0.2$ and that in $\rho'_*(x)$ at $x = 1-\eta = 0.8$ are also verified to satisfy (4.13) and (4.14) respectively. The foregoing form of the invariant density has been completely verified by direct numerical simulations of noisy tent-map dynamics with the prescription of re-injection at the nearest boundary as in (4.2), as shown in figure 3(c). The Lyapunov exponent remains equal to ln 2 in this case, by virtue of the fact that $|f'(x)|$ is a constant and the invariant density is normalized to unity according to (4.12).

5. Generalized dimension D_q

We have commented already on the manner in which the Lyapunov exponent λ is altered (or remains unaltered) for the various noisy maps considered. In the same vein, we may ask how the generalized dimension D_q is altered by the addition of noise. This quantity is defined as [11, 14]

$$D_q = \lim_{\epsilon\to 0} (q-1)^{-1} \ln\left(\sum_{j=1}^{N(\epsilon)} (\mu_j)^q\right) \Big/ \ln \epsilon, \tag{5.1}$$

where $-\infty < q < \infty$. In the present context, $N(\epsilon) = (b-a)/\epsilon$, while the measure of the jth cell is given by

$$\mu_j = \int_{a+(j-1)\epsilon}^{a+j\epsilon} \rho(x)\mathrm{d}x \approx \epsilon\rho(a+(j-\tfrac{1}{2})\epsilon). \tag{5.2}$$

Therefore

$$\sum_{j=1}^{N(\epsilon)} (\mu_j)^q = \epsilon^{q-1} \sum_{j=1}^{N(\epsilon)} \epsilon[\rho(a+(j-\tfrac{1}{2})\epsilon)]^q. \tag{5.3}$$

As the sum on the right-hand side tends to $\int_a^b [\rho(x)]^q \mathrm{d}x$, the behaviour of D_q is governed by the integrability or otherwise of $\rho^q(x)$, i.e., by the nature of the zeros and singularities of $\rho(x)$ [14]. It is at once evident that $D_q = D_0 = 1$ in all cases in which the invariant density is a constant. A simple non-trivial case is once again provided by the symmetric cusp map (3.18), for which $\rho^0(x) = (1-x)/2$ (3.19). Owing to the vanishing of $\rho^0(x)$ at $x = 1$, we have, for $q > -1$,

$$D_q \to \lim_{\epsilon\to 0} (q-1)^{-1} \ln \epsilon^{q-1} / \ln \epsilon = 1, \quad (q > -1). \tag{5.4}$$

For $q \leq -1$, on the other hand, we find

$$D_q \to \lim_{\epsilon\to 0} (q-1)^{-1} \ln \epsilon^{2q} / \ln \epsilon = 2q/(q-1), \quad (q \leq -1) \tag{5.5}$$

since $\mu_N \approx \epsilon\rho(-1 + (N - \frac{1}{2})\epsilon) = \epsilon^2/4$, on setting $a = -1$, $b = 1$, $N\epsilon = b - a = 2$. The question is: how is this result affected by the addition of noise?

In § 3.3, we have seen how the unperturbed density $\rho^0(x) = (1-x)/2$ is modified by various kinds of noise distributions $g(\xi)$ under periodic boundary conditions. For the purpose at hand, the relevant point is that the noise generically eliminates the zero of the density at $x = 1$, without introducing any fresh zeroes or singularities in the interval (cf. (3.24) and (3.26) for $\mu \neq 0$). Therefore D_q remains equal to 1 for all q, in contrast to the noise-free case. The case $\mu = 0$, corresponding to the noise distribution (3.29), is an exception: the corresponding density $\rho(x)$ (3.30) continues to have a simple zero at $x = 1$, so that D_q is given by (5.4) and (5.5), as in the noise-free case. The question can be analyzed for a general noise distribution $g(\xi)$ by examining the representation (3.22) for $\rho(x)$ in the vicinity of $x = 1$.

We turn now to the effect on D_q of the other boundary condition considered in this paper, namely, that of re-injection at the nearest boundary. Again, taking the tent map $f(x) = 1 - |2x - 1|$ in the unit interval as an illustration, we have an invariant density $\rho(x)$ of the form $\alpha\delta(x) + \beta\delta(1-x) + \rho_*(x)$ (cf. (4.16)). Breaking up the interval $[0, 1]$ into N parts with $N\epsilon = 1$, we now have

$$\mu_j = \epsilon\, \rho_*((j - \tfrac{1}{2})\epsilon), \quad 2 \leq j \leq (N-1), \tag{5.6}$$

while

$$\mu_1 = \alpha + \epsilon\rho_*\left(\frac{\epsilon}{2}\right), \quad \mu_N = \beta + \epsilon\rho_*((N - \tfrac{1}{2})\epsilon). \tag{5.7}$$

As α, $\beta > 0$, this leads to

$$D_q \longrightarrow \lim_{\epsilon\to 0} (q-1)^{-1} \frac{\ln [\alpha^q + \beta^q + \epsilon^{q-1} \sum_{j=2}^{N-1} \epsilon\rho_*^q((j - \frac{1}{2})\epsilon)]}{\ln \epsilon}. \tag{5.8}$$

Although $\rho_*(x)$ has a finite discontinuity at $x = \eta$ and cusps at $x = (1+\eta)/2$ and $x = (1-\eta)$ (we consider uniformly distributed noise with maximum amplitude η), $\rho_*^q(x)$ has no zeros or infinities, and is integrable in $[0, 1]$ for all finite q. Therefore, for $q < 1$ the factor ϵ^{q-1} dominates, and $D_q = 1$. On the other hand, this factor tends to zero for $q > 1$, and $D_q = 0$. For $q = 1$, we have

$$D_1 = \lim_{\epsilon\to 0} \left(\sum_{j=1}^{N(\epsilon)} \mu_j \ln \mu_j \right) \Big/ \ln \epsilon. \tag{5.9}$$

Using (5.6) and (5.7), we find

$$D_1 = \lim_{\epsilon \to 0} \sum_{j=1}^{N(\epsilon)} \epsilon\rho_*((j-\tfrac{1}{2})\epsilon) \ln [\epsilon\rho_*((j-\tfrac{1}{2})\epsilon)]/\ln \epsilon, \tag{5.10}$$

which reduces to

$$D_1 = \int_0^1 \mathrm{d}x\, \rho_*(x). \tag{5.11}$$

Collecting these results and using (4.12), we have

$$D_q = \begin{cases} 1, & q < 1, \\ 1-\alpha-\beta, & q = 1, \\ 0, & q > 1. \end{cases} \tag{5.12}$$

The generalized dimensions of the attractor corresponding to the noisy tent map, under the prescription of re-injection at the nearest boundary, are thus modified considerably in comparision with the uniform value $D_q = 1$ for the unperturbed case.

6. Concluding remarks

We have shown that the form of the invariant density for noisy one-dimensional maps in the fully chaotic regime is quite sensitive to the boundary conditions imposed. The latter are necessary to prevent the arbitrary omission or clipping of orbits that overflow the interval when the noise is added. While it is well known that very small amplitude noise helps 'smoothen' the invariant density in general, our results are valid for arbitrary noise amplitude, and show that the invariant density at fully developed chaos can differ significantly from that in the noise-free case. The effect of a marginally stable fixed point can be diminished, as was shown in the case of the cusp map. On the other hand, an initially uniform invariant density can develop considerable structure, as in the case of the tent map with the re-injection prescription used in § 4: δ-function spikes occur at the end-points of the interval, and the unstable fixed point at $x = 0$ is the source of an enhanced boundary layer in its vicinity, as well. Similar effects occur in the case of other maps such as the cusp map and the logistic map. Our results suggest that care must be exercised in drawing conclusions from numerical studies on the aspects of chaotic dynamics considered here, as the combination of external noise and boundary conditions can, and do, lead to non-trivial and sizeable modifications of what one obtains in the absence of noise.

Appendix

We begin with equations (4.5) to (4.7) for α, β and $\rho_*(x)$. Using the rectangular distribution (3.23) for $g(\xi)$ and setting $f(a) = f(b) = a$, (4.5) for α yields

$$\alpha = \tfrac{1}{2}(\alpha+\beta) + \frac{1}{2\eta}\int_{-\eta}^{\eta} \mathrm{d}\xi \int_a^b \mathrm{d}y\, \theta(a - f(y) - \xi)\, \rho_*(y). \tag{A1}$$

The step function is non-zero only for $f(y)+\xi<a$, since $f(y)\geq a$, this is possible only for $\xi<0$. Now the inequality $f(y)<z$ where $z\in[a,b]$ implies that $y\in[a,y_1(z)]\cup[y_2(z),b]$. Therefore

$$\alpha=\tfrac{1}{2}(\alpha+\beta)+\frac{1}{2\eta}\int_{-\eta}^{0}\mathrm{d}\xi\left[\int_{a}^{y_1(a-\xi)}\mathrm{d}y\,\rho_*(y)+\int_{y_2(a-\xi)}^{b}\mathrm{d}y\,\rho_*(y)\right]. \tag{A2}$$

Changing variables to $-\xi$ and interchanging the order of integration, we have

$$\alpha=\beta+\frac{1}{\eta}\int_{a}^{y_1(a+\eta)}\mathrm{d}y\,\rho_*(y)\int_{f_1(y)-a}^{\eta}\mathrm{d}\xi+\frac{1}{\eta}\int_{y_2(a+\eta)}^{b}\mathrm{d}y\,\rho_*(y)\int_{f_2(y)-a}^{\eta}\mathrm{d}\xi, \tag{A3}$$

where f_1 and f_2 are, respectively, the ascending and descending branches of f. Equation (4.8) for α follows immediately. The appropriate branch of $f(y)$ in the integrand of that equation is automatically selected by the range of integration.

Similarly, (4.6) yields

$$\beta=\frac{1}{2\eta}\int_{-\eta}^{\eta}\mathrm{d}\xi\left[(\alpha+\beta)\theta(a-b+\xi)+\int_{a}^{b}\mathrm{d}y\,\theta(f(y)+\xi-b)\rho_*(y)\right]. \tag{A4}$$

As η is certainly less than $b-a$, the first term on the right-hand side does not contribute anything. The second term reduces to

$$\beta=\frac{1}{2\eta}\int_{0}^{\eta}\mathrm{d}\xi\int_{y_1(b-\xi)}^{y_2(b-\xi)}\mathrm{d}y\,\rho_*(y). \tag{A5}$$

Interchanging the order of integration,

$$\beta=\frac{1}{2\eta}\left[\int_{y_1(b-\eta)}^{y_m}\mathrm{d}y\,\rho_*(y)\int_{b-f_1(y)}^{\eta}\mathrm{d}\xi+\int_{y_m}^{y_2(b-\eta)}\mathrm{d}y\,\rho_*(y)\int_{b-f_2(y)}^{\eta}\mathrm{d}\xi\right], \tag{A6}$$

where the peak in $f(y)$ occurs at $y=y_m$. Equation (4.9) for β follows immediately.

Finally, (4.7) becomes

$$\rho_*(x)=\frac{1}{2\eta}\int_{-\eta}^{\eta}\mathrm{d}\xi\left[(\alpha+\beta)\delta(x-a-\xi)+\int_{a}^{b}\mathrm{d}y\,\delta(x-f(y)-\xi)\rho_*(y)\right]. \tag{A7}$$

The first term in the integrand only contributes if x lies between a and $a+\eta$. Equation (4.10) therefore follows.

Acknowledgments

This work was commenced during a visit by VB to Université Libre de Bruxelles, the research being supported in part by the Poles d'Attraction Interuniversitaires and Actions de Recherche Concertées programmes, Belgium. VB is deeply indebted to Suresh Govindarajan for discussions and for his invaluable help with the numerical work including the figures, and in the preparation of the manuscript.

References

[1] J P Crutchfield, J D Farmer and B A Huberman, *Phys. Rep.* **92**, 45 (1982)
[2] R Kapral, M Schell and S Fraser, *J. Phys. Chem.* **86**, 2205 (1982)
[3] A Lasota and M C Mackey, *Physica* **D28**, 143 (1987)
[4] E Knobloch and J B Weiss, in *Noise in nonlinear dynamical systems* edited by F Moss and P V E McClintock (Cambridge Univ. Press, Cambridge, 1989) vol. 2, p. 65
[5] P Talkner and P Hänggi, in *Noise in nonlinear dynamical systems* edited by F Moss and P V E McClintock (Cambridge Univ. Press, Cambridge, 1989) vol. 2, p. 87
[6] S P Hirshman and J C Whitson, *Phys. Fluids* **25**, 967 (1982)
[7] A B Rechester and R B White, *Phys. Rev.* **A27**, 1203 (1983)
[8] R Graham, *Phys. Rev.* **A28**, 1679 (1983)
[9] P Reimann and P Talkner, *Phys. Rev.* **A44**, 6348 (1991)
[10] A Lasota and M C Mackey, *Chaos, fractals and noise* (Springer-Verlag, New York, 1994)
[11] S N Rasband, *Chaotic dynamics of nonlinear systems* (Wiley, New York, 1989)
[12] P C Hemmer, *J. Phys.* **A17**, L247 (1984)
[13] H Haken and G Mayer-Kress, *Z. Phys.* **B43**, 185 (1981)
[14] E Ott, *Chaos in dynamical systems* (Cambridge Univ. Press, Cambridge, 1993)

physics pp. 129–142

Symplectic integration of nonlinear Hamiltonian systems

GOVINDAN RANGARAJAN
Department of Mathematics and Centre for Theoretical Studies, Indian Institute of Science, Bangalore 560 012, India

Abstract. There exist several standard numerical methods for integrating ordinary differential equations. However, if one is interested in integration of Hamiltonian systems, these methods can lead to wrong results. This is due to the fact that these methods do not explicitly preserve the so-called 'symplectic condition' (that needs to be satisfied for Hamiltonian systems) at every integration step. In this paper, we look at various methods for integration that preserve the symplectic condition.

Keywords. Symplectic integration; hamiltonian systems.

PACS Nos **02.20; 03.20**

1. Introduction

In this paper, we consider numerical integration methods for Hamiltonian systems. In particular, we would be interested in long-term integration of these systems. In these cases, it is important to preserve the Hamiltonian nature of the system at every integration step. Otherwise, one can get spurious damping or even chaotic behaviour which is not present in the original system. Such behaviour can obviously lead to wrong predictions regarding the long-term stability of the Hamiltonian system being studied.

In this paper, we look at various integration methods that overcome the above problem [1–20]. Such integration methods go by the name of symplectic integrators. In § 2, we introduce the basic concepts. In § 3, we consider the generating function methods for symplectic integration. Section 4 is devoted to symplectic Runge–Kutta (RK) and Runge–Kutta–Nystrom (RKN) methods. Section 5 introduces the Lie algebraic methods for integrating Hamiltonian systems. A better formulation of the generating function methods in this language is given. The jolt map factorization method for symplectic integration is also considered in some detail. Finally, integration using solvable maps is discussed. The concluding remarks can be found in § 6.

2. Basic concepts

Consider the following set of $2n$ first order differential equations:

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\partial H(q,p)}{\partial p}, \quad \frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial H(q,p)}{\partial q}, \tag{2.1}$$

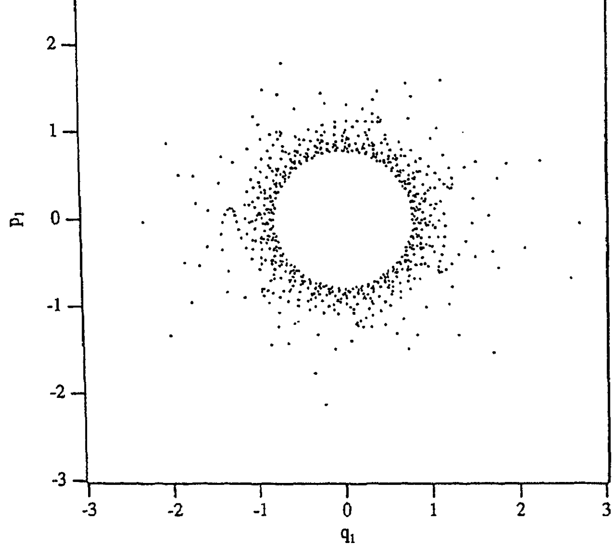


Figure 1. Numerical integration results using a non-symplectic fourth-order Taylor series map for the anharmonic oscillator. The initial condition used is $(q_1(0), p_1(0)) = (0.8, 0.0)$.

where q, $p \in R^n$. The initial conditions are given by $q(t_0) = q_0$ and $p(t_0) = p_0$. The above system of differential equations defines a Hamiltonian system where the Hamiltonian function H is a function of the variables q_i, p_i $(i = 1, 2, \ldots, n)$. The variables q_i and p_i constitute the phase space of the Hamiltonian system. Typically, q_i's are the (generalized) coordinates and p_i's the (generalized) momenta of the system.

Consider the Poisson bracket [,] of two phase space functions $f(q,p)$ and $g(q,p)$ defined as

$$[f(q,p), g(q,p)] = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i}\right]. \tag{2.2}$$

Hamiltonian systems possess the remarkable property that they preserve the fundamental Poisson bracket $[q_i, p_j]$, that is

$$[q_i(t), p_j(t)] = [q_i(0), p_j(0)], \quad \forall\, i, j. \tag{2.3}$$

Equivalently, Hamiltonian systems preserve the symplectic 2-form $\mathrm{d}p \wedge \mathrm{d}q$. This condition is called the symplectic condition.

In general H is a complicated function of q's and p's. Consequently, the resultant equations of motion [cf. eq. (2.1)] are nonlinear ordinary differential equations. Generically, such systems of nonlinear equations do not admit an analytic solution, that is, the system is non-integrable. Hence, one is forced to integrate these sets of ordinary differential equations numerically. There exist several standard methods for

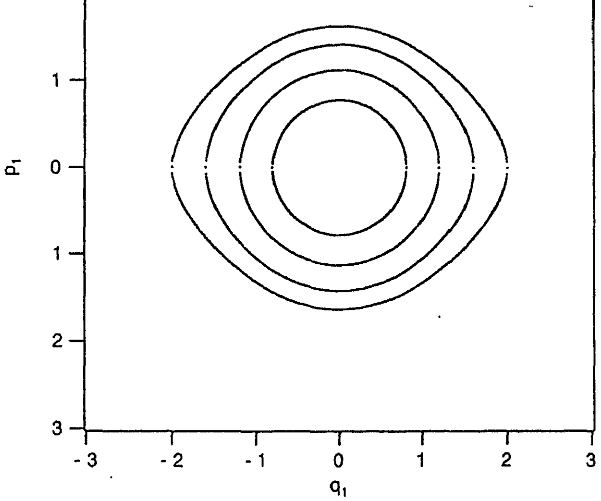


Figure 2. Numerical integration results using the exact symplectic map for the anharmonic oscillator. The initial conditions used are $(q_1(0), p_1(0)) = (0.8, 0.0)$, $(1.2, 0.0)$, $(1.6, 0.0)$, and $(2.0, 0.0)$.

numerical integration of a system of first order ordinary differential equations. However, these are general purpose methods and are not geared explicitly towards Hamiltonian systems. In particular, they do not preserve the symplectic condition at every step. For short-term integration this does not lead to much problems. For long-term integration, these non-symplectic methods can be a disaster. Since the symplectic condition is not preserved, even the qualitative nature of the solutions obtained by these methods can be very different from that of the exact solution. For example, one can get spurious damping or chaos where there is none. This can lead to wrong deductions regarding the long-term stability of the Hamiltonian system in question.

An example of how things can go wrong is shown in figure 1. This figure gives the result of integrating the nonlinear Hamiltonian system defined by the following Hamiltonian using a non-symplectic method

$$H = \frac{q_1^2 + p_1^2}{2} - \frac{q_1^4}{24}. \tag{2.4}$$

Looking at the figure, one might conclude that chaotic behaviour is present in the system. However, the exact solution of the system exhibited in figure 2 shows that this is not true. One way to avoid such a situation while using non-symplectic methods is to reduce the step size so that the symplectic violation is very small at each step. However, the computational cost of such a procedure for long-term numerical integration would be prohibitive. Therefore, one is lead to look for integration schemes which explicitly preserve the symplectic condition at every step. Such schemes are called symplectic integration methods and we look at some of them in the following sections.

3. Generating function methods

Symplectic integration methods were first discovered by DeVogelaere [1]. However, these results were not widely known. In 1983, Ruth [2] independently discovered symplectic integration methods using canonical generating functions.

The basic idea behind this method is as follows. For simplicity, let us consider a Hamiltonian given by the special form

$$H(q,p) = A(p) + V(q). \tag{3.1}$$

Ideally, we would like to make a canonical transformation such that the new Hamiltonian H' in the new variables is identically zero. However, this is not practical. Hence, we attempt to make H' zero up to a given order t^k [2, 3]

$$H'(q',p',t) = O(t^{k+1}), \tag{3.2}$$

where q' and p' are the new variables after the canonical transformation. Using the equations of motion [cf. eq. (2.1)], we get

$$q'(t) = q_0 + O(t^{k+1}), \quad p'(t) = p_0 + O(t^{k+1}). \tag{3.3}$$

Thus, the motion is simple in the new coordinates. Once this is accomplished, we can invert the canonical transformation to get the motion in the original variables (accurate to order k). Therefore we get a map from $q(0), p(0)$ to $q(t), p(t)$ (again accurate to order k). Now we can repeat the process to go from $q(t), p(t)$ to $q(2t), p(2t)$ and so on. Thus, we have an integration algorithm (of order k) for the Hamiltonian system which allows us to step forward in time starting from the intitial conditions. Of course, we will typically take the time step t at every stage to be small so that the error generated (which is of order t^{k+1}) is small. Since canonical transformations explicitly preserve the symplectic condition, we in fact have a symplectic integrator of order k.

We did not specify above how the canonical transformation is to be performed. This is done using the generating function method [21]. We use a generating function of the new coordinates and the old momenta [2, 3]

$$F_3(q',p,t) = -q' \cdot p + G(q',p,t). \tag{3.4}$$

Now we derive a simple symplectic integration method using the technique described above. Let the generating function be specified by [2]

$$G = -H(q',p)t = -[A(p) + V(q')]t. \tag{3.5}$$

Using the canonical transformation equations appropriate for this generating function [21] we get

$$\begin{aligned} p_i' &= p_i + \frac{\partial V(q')}{\partial q_i'}t, \\ q_i &= q_i' + \frac{\partial A(p)}{\partial p_i}t. \end{aligned} \tag{3.6}$$

Note that the momentum equation would in general be an implicit equation in p which has to be inverted to obtain p as a function of p'. This can be done explicitly only for terms of q' and p' as follows [2]:

$$H' \sim t \sum_{i=1}^{n} \frac{\partial V(q')}{\partial q'_i} \frac{\partial A(p')}{\partial p'_i} + O(t^2). \tag{3.7}$$

Consequently, one gets

$$q' = q_0 + O(t^2), \quad p' = p_0 + O(t^2). \tag{3.8}$$

Thus, we have succeeded in obtaining a symplectic integration method of order 1 for our simple Hamiltonian.

Following the same procedure, one can in principle get symplectic integration methods of higher order. However, it is not easy to obtain the necessary canonical transformations. In general, one uses a series of transformations to attain the final goal. For example, for the special Hamiltonian given in (3.1), a fourth order symplectic integration method is given by the following equations [3]

$$p_i^{(l+1)} = p_i^{(l)} - c_l\, t \frac{\partial V}{\partial q_i^{(l)}}, \quad q_i^{(l+1)} = q_i^{(l)} - d_l\, t \frac{\partial A}{\partial p_i^{(l+1)}}; \quad l = 0,1,2,3,4. \tag{3.9}$$

Here the index l labels successive canonical transformations. Moreover, $q^{(0)} = q_0$ and $p^{(0)} = p_0$ (the initial conditions). The unknown coefficients c_l, d_l are determined by the condition that the new Hamiltonian in terms of the final canonical variables is zero upto order t^4. One possible solution is given below [3, 5]:

$$c_1 = c_4 = a + \tfrac{1}{2},\ c_2 = c_3 = -a,\ d_1 = d_3 = 2a + 1,\ d_2 = -4a - 1,\ d_4 = 0, \tag{3.10}$$

where $a = 0.1756\ldots$.

The necessary canonical transformations become more and more difficult to solve for as one goes to higher orders. Further, for complicated Hamiltonians the inversion of the momenta equations becomes impossible analytically. One has to invert them numerically (typically using Newton–Raphson method) with its attendent convergence problems. Also, this slows down the integration algorithm considerably. We shall see that these problems are solved by using Lie algebraic methods.

4. Symplectic RK and RKN methods

In this section, we discuss Runge–Kutta and Runge–Kutta–Nystrom methods that are explicitly symplectic [14–20]. We start with the set of $2n$ first-order differential equations for a Hamiltonian system [cf. eq. (2.1)] which we reformulate as follows:

$$\frac{dz}{dt} = f(z), \quad z(t_0) = z_0,$$

where $z = (q, p)$ is a $2n$-vector and f is a vector function specified by the right hand side of (2.1).

An s-stage Runge–Kutta method to integrate the above system of differential equations is given by [20]

$$Z_i = z_n + h \sum_{j=1}^{s} a_{ij} f(Z_j), \quad i = 1, 2, \ldots, s, \tag{4.1}$$

$$z_{n+1} = z_n + h \sum_{j=1}^{s} b_j f(Z_j). \tag{4.2}$$

Here h is the step size and z_n, z_{n+1} are the values of phase space variables at the end of the nth and $(n+1)$th integration steps respectively. The coefficients a_{ij} and b_j have specific values depending on the RK method under consideration. The RK method is said to be explicit [20] if a_{ij} vanishes for $i \leq j$. In this case, the vectors Z_j can be solved recursively without having to solve any implicit equations. The method is called diagonally implicit if a_{ij} vanishes for $i < j$ and implicit otherwise. It can be shown [17] that only implicit RK methods can be symplectic. These are symplectic if the following condition is satisfied [17]:

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \quad i, j = 1, 2, \ldots, s.$$

An example of a two-stage implicit RK method which is symplectic is specified by

$$a_{11} = a_{22} = \frac{1}{4}, \quad a_{12} = \frac{1}{4} - \frac{\sqrt{3}}{6}, \quad a_{21} = \frac{1}{4} + \frac{\sqrt{3}}{6}, \quad b_1 = b_2 = \frac{1}{2}. \tag{4.3}$$

This can be shown to be a symplectic integrator of order 4. In fact, this is one member of a family of symplectic RK methods which go by the name of Gauss–Legendre methods [20]. However, these are computationally costly methods. Iserles[16] has derived more efficient symplectic RK methods.

Next, we consider Runge–Kutta–Nystrom (RKN) methods. For simplicity, we restrict ourselves to Hamiltonians of the form $H = p^2/2 + V(q)$. The equations of motion for such Hamiltonians can be written as

$$\frac{d^2 q}{dt^2} = f(q),$$

where f is the gradient of $-V$. The above system can be solved by an s-stage RKN method [20]

$$Q_i = q_n + hc_i p_n + h^2 \sum_{i=1}^{s} a_{ij} f(Q_j), \quad j = 1, 2, \ldots, s, \tag{4.4}$$

$$p_{n+1} = p_n + h \sum_{j=1}^{s} b_j f(Q_j), \tag{4.5}$$

$$q_{n+1} = q_n + hp_n + h^2 \sum_{j=1}^{s} \bar{b}_j f(Q_j). \tag{4.6}$$

This method is symplectic if the following conditions are satisfied [15,18]

$$\bar{b}_i - b_i + b_i c_i = 0, \quad i = 1, 2, \ldots, s, \tag{4.7}$$

$$b_i a_{ij} - b_j a_{ji} - b_i \bar{b}_j + b_j \bar{b}_i = 0, \quad i = 1, 2, \ldots, j-1; \quad j = 2, 3, \ldots, s. \tag{4.8}$$

Unlike RK methods, explicit RKN methods can also be symplectic.

The simplest symplectic one-stage RKN method is the Stormer–Verlet method of order 2 specified by

$$a_{11} = 0, \quad c_1 = 0.5, \quad b_1 = 0.5, \quad \bar{b}_1 = 1.$$

The above method is used extensively in molecular dynamics simulations. Higher order symplectic RKN methods have been derived. The highest order method currently available is the explicit 13-stage method of order 8 obtained by Calvo and Sanz–Serna [19]. Symplectic RKN methods have also been obtained using collocation and perturbed collocation techniques [20].

5. Lie algebraic methods

One of the approaches to symplectic integration is the use of Lie algebraic methods [3, 5, 7–13]. Before going into the details we need to consider some mathematical preliminaries. The common denominator in all these approaches is the representation of Hamiltonian systems by symplectic maps [22] which are then iterated to obtain the time evolution of the system.

We start by defining certain mathematical objects. For simplicity, we restrict ourselves to three degree-of-freedom systems. Let us denote the collection of six phase-space variables q_i, p_i $(i = 1, 2, 3)$ by the symbol z:

$$z = (q_1, p_1, q_2, p_2, q_3, p_3). \tag{5.1}$$

The Lie operator corresponding to a phase-space function $f(z)$ is denoted by $:f(z):$. It is defined by its action on a phase-space function $g(z)$ as shown below

$$:f(z):g(z) = [f(z), g(z)]. \tag{5.2}$$

Here $[f(z), g(z)]$ denotes the usual Poisson bracket of the functions $f(z)$ and $g(z)$ [cf. eq. (2.2)]. Next, we define the exponential of a Lie operator. It is called a Lie transformation and is given by

$$\mathrm{e}^{:f(z):} = \sum_{n=0}^{\infty} \frac{:f(z):^n}{n!}. \tag{5.3}$$

Powers of $:f(z):$ that appear in the above equation are defined recursively by the relation

$$:f(z):^n g(z) = :f(z):^{n-1} [f(z), g(z)], \tag{5.4}$$

with

$$:f(z):^0 g(z) = g(z). \tag{5.5}$$

For further details regarding Lie operators and Lie transformations, see [22].

$\mathcal{M}$ [22] as follows:

$$\frac{\mathrm{d}\mathcal{M}}{\mathrm{d}t} = \mathcal{M}\colon -H(z_0)\colon. \tag{5.6}$$

Symplectic maps are maps whose Jacobian matrices $M(z)$ satisfy the following symplectic condition

$$\widetilde{M(z)}JM(z) = J\,, \tag{5.7}$$

where $\tilde{M}$ is the transpose of M and J is an antisymmetric matrix defined as follows:

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{pmatrix}. \tag{5.8}$$

Matrices M satisfying (5.7) are called symplectic matrices and the corresponding maps $\mathcal{M}$ symplectic maps. Symplectic maps explicitly preserve the symplectic condition. It can be shown [22] that the set of all $\mathcal{M}$'s forms an infinite dimensional Lie group of symplectic maps. On the other hand, the set of all real 6×6 symplectic matrices forms the finite dimensional real symplectic group $Sp(6, R)$.

The above map $\mathcal{M}$ gives the final state $z^{(1)}$ of a particle after one time step as a function of its initial state $z^{(0)}$:

$$z^{(1)} = \mathcal{M}z^{(0)}. \tag{5.9}$$

To obtain the state of a particle after N time steps, one has to merely iterate the above mapping N times i.e.

$$z^{(N)} = \mathcal{M}^N z^{(0)}. \tag{5.10}$$

Thus, we have obtained an integrator that is symplectic (since $\mathcal{M}$ preserves symplecticity). Moreover, we have the values of phase space variables after one time step as explicit functions of the initial values.

5.1 *Generating function methods*

We reformulate the generating function method using Lie algebraic techniques [3]. For a time-independent Hamiltonian, (5.6) can be solved formally to give

$$\mathcal{M}(t) = \exp(\colon -tH(z_0)\colon). \tag{5.11}$$

If one could solve this explicitly, one would get a symplectic integrator valid to arbitrary order since symplectic maps explicitly preserve the symplectic condition. However, unless the Hamiltonian is integrable, one can not evaluate the Lie transformation on the right hand side. On the other hand, for special Hamiltonians, it may be possible to split

$$\begin{aligned} &H = H_1 + H_2, \\ &\mathcal{N}_i(t) \equiv \exp(:-tH_i:) \quad i = 1, 2, \end{aligned} \tag{5.12}$$

where the symplectic map $\mathcal{N}_i$ can be evaluated exactly (or up to order t^k, the order of the symplectic integration method that we wish to derive). The above restriction is a drawback of this method (equivalent to the invertibilty of momenta equation in § 3). We will overcome this restriction in the next subsection.

Under the assumption that H can be split as above, one can construct symplectic integration methods as follows. Consider

$$\mathcal{N}(t) = \mathcal{N}_1(t)\mathcal{N}_2(t). \tag{5.13}$$

Since each symplectic map $\mathcal{N}_i$ is a canonical transformation [22], $\mathcal{N}$ is a product of two canonical transformations. Therefore, this is just a reformulation of the generating function method considered in § 3. Using the group-theoretical Campbell–Baker–Hausdorff (CBH) formula [23], one obtains

$$\mathcal{N}(t) = \mathcal{N}_1(t)\mathcal{N}_2(t) = \exp(:-tH_1 - tH_2 - t^2[H_1, H_2]/2 + \cdots:). \tag{5.14}$$

Using eq. (5.12), we get

$$\mathcal{N}(t) = \mathcal{M}(t) + O(t^2). \tag{5.15}$$

Thus, $\mathcal{N}(t)$ is a symplectic first-order approximation to $\mathcal{M}(t)$ and consequently, we get $z(t)$ as a function of $z(0)$ accurate to order t. In other words, we have succeeded in obtaining a first-order symplectic integrator. If we apply this to the H given in (3.1) with $H_1 = A(p)$ and $H_2 = V(q)$, we get back (3.6). This again demonstrates the equivalence of this method with the generating function method described in § 3.

To obtain a fourth-order symplectic integrator, we start with the relation [3]

$$\mathcal{N}(t) = e^A e^B e^{\alpha A} e^{\beta B} e^{\alpha A} e^B e^A,$$

where

$$2(1+\alpha)A = :-tH_1:, \quad (2+\beta) = :-tH_2:.$$

If

$$\alpha = 1 - 2^{1/3}, \quad \beta = \frac{4\alpha^2 - 4\alpha - 2}{(\alpha+1)^2},$$

we get [3]

$$\mathcal{N}(t) = \mathcal{M}(t) + O(t^5).$$

This is of course a symplectic integrator of order 4. This is actually the same integrator that we discussed in § 3 in a different form.

The above method was further generalized by Yoshida [7] to obtain symplectic integrators of arbitrary even orders for Hamiltonians of the form given in (3.1).

5.2 *Symplectic completion of symplectic jets*

In this approach [8–10, 12, 13], symplecticity is achieved by refactorizing the truncated symplectic map using the so-called 'jolt maps'.

We start by factorizing the symplectic map $\mathcal{M}$ [cf. eq. (5.6)] representing the Hamiltonian system using the Dragt–Finn factorization theorem [22, 24] from Lie perturbation theory:

$$\mathcal{M} = \hat{M} e^{:f_3:} e^{:f_4:} \cdots e^{:f_n:} \cdots . \tag{5.16}$$

Here $\hat{M}$ gives the linear part of the map and hence has an equivalent representation in terms of the Jacobian matrix $M(0)$ of the map $\mathcal{M}$ at the origin [22]:

$$\hat{M} z_i = M_{ij} z_j = (Mz)_i . \tag{5.17}$$

Thus, $\hat{M}$ is said to be the Lie transformation corresponding to the 6×6 matrix M belonging to $Sp(6, R)$. The infinite product of Lie transformations $\exp(:f_n:)$ $(n = 3, 4, \ldots)$ in (5.16) represents the nonlinear part of $\mathcal{M}$. Here $f_n(z)$ denotes a homogeneous polynomial (in z) of degree n uniquely determined by the factorization theorem.

One can not use $\mathcal{M}$ in the form given in (5.16) for numerical computations since it involves an infinite number of Lie transformations. Using the perturbation theory approach, we therefore truncate $\mathcal{M}$ as follows:

$$\mathcal{M} \approx \hat{M} e^{:f_3:} e^{:f_4:} \cdots e^{:f_P:} . \tag{5.18}$$

However, each exponential $e^{:f_n:}$ in $\mathcal{M}$ still contains an infinite number of terms in its Taylor series expansion. One possible way around this is to truncate the Taylor series generated by the Lie transformations to order P giving the following truncated map (denoted by $\mathcal{M}_P$):

$$\begin{aligned} \mathcal{M}_P z = M(1 + :f_3: + \cdots)(1 + :f_4: + \cdots) \cdots \\ \cdots (1 + :f_p: + \cdots) z . \end{aligned} \tag{5.19}$$

Here the power series is truncated in such a way that the highest order term generated is z^{P-1}. The truncated map $\mathcal{M}_P$ is called a symplectic jet of order P (for a more precise definition, see [10]). Despite its name, $\mathcal{M}_P$ is not symplectic because of the truncation of the Taylor series. Therefore, it can lead to spurious damping or growth (as discussed earlier).

The basic idea behind the approaches followed [8, 10, 12] is to refactorize $\mathcal{M}_P$ [cf. eq. (5.19)] as a product of symplectic maps that can be evaluated exactly. This process of refactorizing $\mathcal{M}_P$ goes by the name of 'symplectic completion of symplectic jets'.

We begin this process by defining jolt maps. Consider the symplectic map given by $e^{:g(z):}$ where $g(z)$ is a function of the phase space variables z. It is called a jolt map if $:g(z):$ is a nilpotent operator of rank 2 i.e. if the following condition is satisfied

$$:g(z):^2 z = 0 . \tag{5.20}$$

The function $g(z)$ is then called a jolt function. We note that jolt maps have only two non-zero terms in their Taylor series expansions [cf. eq. (5.3)]. Hence they can be evaluated exactly without any truncation. The term jolt map was first introduced in [12]. An example of jolt map [10] is given below:

Here $f(q_1, q_2, q_3)$ is an nth degree polynomial in variables q_1, q_2, and q_3 and $\hat{R}$ is the Lie transformation corresponding to a 6×6 matrix R belonging to any subgroup of $Sp(6, R)$. $\hat{R}$ is given by the following relation [cf. eq. (5.17)]

$$\hat{R} z_i \; = \; R_{ij} z_j \; = \; (Rz)_i. \tag{5.22}$$

Now we can formulate the problem [8]. Given the map $\mathcal{M}_P$, find another map $\mathcal{J}$ specified by the following product of K jolt maps

$$\mathcal{J} = \hat{M} \mathrm{e}^{:g_3^{(1)}+g_4^{(1)}+\cdots+g_P^{(1)}:} \mathrm{e}^{:g_3^{(2)}+g_4^{(2)}+\cdots+g_P^{(2)}:} \ldots \mathrm{e}^{:g_3^{(K)}+g_4^{(K)}+\cdots+g_P^{(K)}:} \tag{5.23}$$

such that this map agrees with $\mathcal{M}_P$ to order P i.e.

$$\mathcal{J} \cong \mathcal{M}_P \quad \text{to order } P. \tag{5.24}$$

Here $g_n^{(i)}$'s are (homogeneous) jolt polynomials of degree n given by the following relation

$$g_n^{(i)} = \beta_n^{(i)} \hat{R}_i q_1^n \quad i = 1, 2, \ldots, K, \tag{5.25}$$

where $\beta_n^{(i)}$ is a real coefficient. The matrices R_i belong to a subgroup of $Sp(6, R)$ (including $Sp(6, R)$ itself) and $\hat{R}_i$ denotes the Lie transformation corresponding to these matrices [cf. eq. (5.22)]. Since $\mathcal{J}$ is given as a product of finite number of jolt maps, it can be evaluated exactly without any truncation. Thus it gives a symplectic approximation to $\mathcal{M}_P$ accurate to order P.

By comparing terms of same order in $\mathcal{M}_P$ and $\mathcal{J}$, the problem of obtaining a symplectic completion for $\mathcal{M}_P$ can be reduced to the following general problem: Given a nth degree homogeneous polynomial f_n and K matrices R_i, find the coefficients $\beta_n^{(i)}$'s such that the following condition is satisfied

$$\sum_{i=1}^{K} \beta_n^{(i)} \hat{R}_i q_1^n = f_n. \tag{5.26}$$

Typically, one has more $\beta_n^{(i)}$'s than there are equations. So, one imposes constraints on these coefficients. In [10], the sum of squares of these coefficients is minimized. A more sophisticated approach can be found in [12, 13].

To proceed further, following the approach given in [10], one goes to the continuum limit of the above problem. We get the following continuum problem: Given an nth degree homogeneous polynomial f_n and a subgroup G of $Sp(6, R)$ on which invariant integration is well defined, find the function $g(u)$ such that the following condition is satisfied

$$f_n = \int_G \mathrm{d}u \, g(u) \hat{R}(u) q_1^n. \tag{5.27}$$

Here u denotes a general element of the group G and $\hat{R}(u)$ denotes the Lie transformation corresponding to u. All integrations are invariant integrations performed over the group G. Again, an appropriate constraint on $g(u)$ is imposed.

This problem has been solved [10] by taking the group G to be $SU(3)$. The solution is given by:

$$g(u) = \sum_{j,m} \frac{\mathrm{d}\phi_m^j}{\xi^j} \, \bar{\mathcal{D}}^j_{mm_j}(u), \quad j \leq n. \tag{5.28}$$

Here, ξ_j's (all of which can be shown to be non-zero) and ϕ_m^j's are the coefficients

$$\mathcal{M} = \hat{M} e^{:f_3:} e^{:f_4:} \cdots e^{:f_n:} \cdots . \tag{5.16}$$

Here $\hat{M}$ gives the linear part of the map and hence has an equivalent representation in terms of the Jacobian matrix $M(0)$ of the map $\mathcal{M}$ at the origin [22]:

$$\hat{M} z_i = M_{ij} z_j = (Mz)_i . \tag{5.17}$$

Thus, $\hat{M}$ is said to be the Lie transformation corresponding to the 6×6 matrix M belonging to $Sp(6,R)$. The infinite product of Lie transformations $\exp(:f_n:)$ $(n = 3, 4, \ldots)$ in (5.16) represents the nonlinear part of $\mathcal{M}$. Here $f_n(z)$ denotes a homogeneous polynomial (in z) of degree n uniquely determined by the factorization theorem.

One can not use $\mathcal{M}$ in the form given in (5.16) for numerical computations since it involves an infinite number of Lie transformations. Using the perturbation theory approach, we therefore truncate $\mathcal{M}$ as follows:

$$\mathcal{M} \approx \hat{M} e^{:f_3:} e^{:f_4:} \cdots e^{:f_P:} . \tag{5.18}$$

However, each exponential $e^{:f_n:}$ in $\mathcal{M}$ still contains an infinite number of terms in its Taylor series expansion. One possible way around this is to truncate the Taylor series generated by the Lie transformations to order P giving the following truncated map (denoted by $\mathcal{M}_P$):

$$\begin{aligned} \mathcal{M}_P z = M(1 + :f_3: + \cdots)(1 + :f_4: + \cdots) \cdots \\ \cdots (1 + :f_p: + \cdots) z. \end{aligned} \tag{5.19}$$

Here the power series is truncated in such a way that the highest order term generated is z^{P-1}. The truncated map $\mathcal{M}_P$ is called a symplectic jet of order P (for a more precise definition, see [10]). Despite its name, $\mathcal{M}_P$ is not symplectic because of the truncation of the Taylor series. Therefore, it can lead to spurious damping or growth (as discussed earlier).

The basic idea behind the approaches followed [8, 10, 12] is to refactorize $\mathcal{M}_P$ [cf. eq. (5.19)] as a product of symplectic maps that can be evaluated exactly. This process of refactorizing $\mathcal{M}_P$ goes by the name of ‘symplectic completion of symplectic jets’.

We begin this process by defining jolt maps. Consider the symplectic map given by $e^{:g(z):}$ where $g(z)$ is a function of the phase space variables z. It is called a jolt map if $:g(z):$ is a nilpotent operator of rank 2 i.e. if the following condition is satisfied

$$:g(z):^2 z = 0. \tag{5.20}$$

The function $g(z)$ is then called a jolt function. We note that jolt maps have only two non-zero terms in their Taylor series expansions [cf. eq. (5.3)]. Hence they can be evaluated exactly without any truncation. The term jolt map was first introduced in [12]. An example of jolt map [10] is given below:

$$e^{:\hat{R} f(q_1, q_2, q_3):} . \tag{5.21}$$

Here $f(q_1, q_2, q_3)$ is an nth degree polynomial in variables q_1, q_2, and q_3 and $\hat{R}$ is the Lie transformation corresponding to a 6×6 matrix R belonging to any subgroup of $Sp(6, R)$. $\hat{R}$ is given by the following relation [cf. eq. (5.17)]

$$\hat{R} z_i \;=\; R_{ij} z_j \;=\; (Rz)_i. \tag{5.22}$$

Now we can formulate the problem [8]. Given the map $\mathcal{M}_P$, find another map $\mathcal{J}$ specified by the following product of K jolt maps

$$\mathcal{J} = \hat{M} \mathrm{e}^{:g_3^{(1)}+g_4^{(1)}+\cdots+g_P^{(1)}:} \mathrm{e}^{:g_3^{(2)}+g_4^{(2)}+\cdots+g_P^{(2)}:} \ldots \mathrm{e}^{:g_3^{(K)}+g_4^{(K)}+\cdots+g_P^{(K)}:} \tag{5.23}$$

such that this map agrees with $\mathcal{M}_P$ to order P i.e.

$$\mathcal{J} \cong \mathcal{M}_P \quad \text{to order } P. \tag{5.24}$$

Here $g_n^{(i)}$'s are (homogeneous) jolt polynomials of degree n given by the following relation

$$g_n^{(i)} = \beta_n^{(i)} \hat{R}_i q_1^n \quad i = 1, 2, \ldots, K, \tag{5.25}$$

where $\beta_n^{(i)}$ is a real coefficient. The matrices R_i belong to a subgroup of $Sp(6, R)$ (including $Sp(6, R)$ itself) and $\hat{R}_i$ denotes the Lie transformation corresponding to these matrices [cf. eq. (5.22)]. Since $\mathcal{J}$ is given as a product of finite number of jolt maps, it can be evaluated exactly without any truncation. Thus it gives a symplectic approximation to $\mathcal{M}_P$ accurate to order P.

By comparing terms of same order in $\mathcal{M}_P$ and $\mathcal{J}$, the problem of obtaining a symplectic completion for $\mathcal{M}_P$ can be reduced to the following general problem: Given a nth degree homogeneous polynomial f_n and K matrices R_i, find the coefficients $\beta_n^{(i)}$'s such that the following condition is satisfied

$$\sum_{i=1}^{K} \beta_n^{(i)} \hat{R}_i q_1^n = f_n. \tag{5.26}$$

Typically, one has more $\beta_n^{(i)}$'s than there are equations. So, one imposes constraints on these coefficients. In [10], the sum of squares of these coefficients is minimized. A more sophisticated approach can be found in [12, 13].

To proceed further, following the approach given in [10], one goes to the continuum limit of the above problem. We get the following continuum problem: Given an nth degree homogeneous polynomial f_n and a subgroup G of $Sp(6, R)$ on which invariant integration is well defined, find the function $g(u)$ such that the following condition is satisfied

$$f_n = \int_G \mathrm{d}u \, g(u) \hat{R}(u) q_1^n. \tag{5.27}$$

Here u denotes a general element of the group G and $\hat{R}(u)$ denotes the Lie transformation corresponding to u. All integrations are invariant integrations performed over the group G. Again, an appropriate constraint on $g(u)$ is imposed.

This problem has been solved [10] by taking the group G to be $SU(3)$. The solution is given by:

$$g(u) = \sum_{j,m} \frac{\mathrm{d}\phi^j_m}{\xi^j} \bar{\mathcal{D}}^j_{mm_j}(u), \quad j \leq n. \tag{5.28}$$

Here, ξ_j's (all of which can be shown to be non-zero) and ϕ^j_m's are the coefficients obtained by expanding q_1^n and f_n respectively in terms of $SU(3)$ basis vectors. Further, $\bar{\mathcal{D}}^j_{mm_j}(u)$'s are the complex conjugates of the usual D matrices that occur in $SU(3)$ representation theory. Finally, d is the dimension of the $SU(3)$ representation labeled by j.

Once the continuum solution has been obtained, one can go back to the discrete case by using discrete subgroups of $SU(3)$ or by quadrature formulas. See refs [10, 12, 13] for further details.

Now that a solution $g(u)$ has been obtained, we would like to reduce the number of jolt maps involved in the refactorization so that a fast numerical integration method is obtained. A substantial reduction can be achieved by splitting the integration over $SU(3)$ as an integration over $SO(3)$ followed by an integration over the coset space $SU(3)/SO(3)$. In ref. [10] it is shown that the number of jolt maps required by following this approach is less than that required by Irwin [8]. Irwin uses the group $U(1) \times U(1) \times U(1)$. However, there are still many theoretical problems regarding integration over coset spaces etc. which are not fully resolved. Work is still proceeding along these lines.

5.3 *Solvable map method*

In this approach [9,11], the symplectic map $\mathcal{M}$ is refactorized using the so-called 'solvable maps'.

Solvable maps [11] are generalizations of Cremona maps. The class of Cremona maps includes only those symplectic maps for which the Taylor series expansion terminates when acting on phase space coordinates. The class of solvable maps also includes those symplectic maps for which the Taylor series expansion can be summed explicitly. The Lie transformation $\exp(:aq_1^{l+2} + bq_1^{l+1}p_1:)$ is a simple example of a solvable map that is not a kick. One finds that

$$q_1' = \exp(:aq_1^{l+2} + bq_1^{l+1}p_1:)\, q_1 = \frac{q_1}{[1 + lbq_1^l]^{\frac{1}{l}}} \quad \text{for} \quad l \geq 1, \tag{5.29}$$

$$p_1' = \exp(:aq_1^{l+2} + bq_1^{l+1}p_1:)\, p_1 = \frac{E - a(q_1')^{l+2}}{b(q_1')^{l+1}}, \tag{5.30}$$

where

$$E = aq_1^{l+2} + bq_1^{l+1}p_1. \tag{5.31}$$

The basic idea behind the solvable map method is to represent each nonlinear factor $\exp(:f_n:)$ in $\mathcal{M}_P$ [cf. eq. (5.18)] as a product of solvable maps:

$$\exp(:f_n:) = \exp(:g_1:)\, \exp(:g_2:) \cdots \exp(:g_m:) \quad \text{for} \quad n \geq 3. \tag{5.32}$$

For example, consider the Lie representation of a general fourth order map in one dimension [cf. eq. (5.18)]:

$$\mathcal{M}_4 = \hat{M} \exp(:f_3:)\, \exp(:f_4:) \tag{5.33}$$

where

$$f_3 = a_1q_1^3 + a_2q_1^2p_1 + a_3q_1p_1^2 + a_4p_1^3, \tag{5.34}$$

$$f_4 = a_5q_1^4 + a_6q_1^3p_1 + \cdots + a_9p_1^4. \tag{5.35}$$

This can be represented in terms of solvable maps as given below:

$$\begin{aligned}\mathcal{M}_4 =&\hat{M}\exp(:b_1q_1^3+b_2q_1^2p_1:)\exp(:b_3q_1p_1^2+b_4p_1^3:)\\ &\exp(:b_5q_1^4+b_6q_1^3p_1:)\exp(:b_7q_1^2p_1^2:)\exp(:b_8q_1p_1^3+b_9p_1^4:),\end{aligned}\tag{5.36}$$

where

$$b_i=a_i \quad \text{for} \quad i=1,2,\ldots,5 \quad \text{and} \quad i=9,\tag{5.37}$$

$$b_6=a_6-3a_1a_3,\tag{5.38}$$

$$b_7=a_7-9a_1a_4/2-3a_2a_3/2,\tag{5.39}$$

$$b_8=a_8-3a_2a_4.\tag{5.40}$$

Only some preliminary work has been done on this method. It needs a stronger theoretical foundation.

Another related approach is the monomial factorization method. See Refs [25, 26] for further details.

6. Summary

In this paper, we saw that if integration performed on a Hamiltonian system is not symplectic, one can draw wrong conclusions regarding the long-term stability of the system. We looked at various methods for symplectic integration. There is no single ideal method yet. Generating function methods suffer from the drawback that implicit equations have to be solved numerically during the process of integration with its attendent convergence problems. This can also lead to reduction in the speed of the algorithm. Symplectic RK methods are again implicit. Symplectic RKN methods can be explicit but are slow for long-term integration. This is especially so when one is integrating complicated systems like particle storage rings with thousands of components. Integration using symplectic maps can be very fast since one has to merely iterate the maps. Methods using Lie algebraic properties (like jolt factorization method and solvable map method) also give maps where the final values of variables are explicit functions of initial values. For these reasons, these approaches look most promising especially for long-term integration. However, some theoretical problems associated with these methods are not fully resolved.

Acknowledgement

This work was supported in part by the DAE project No. 48/3/96-R&DII/546.

References

[1] R DeVogelaere, *Methods of integration which preserve the contact transformation property of the Hamilton equations*, N7-ONR-43906
[2] R D Ruth, *IEEE Trans. Nucl. Sci.* **30**, 2669 (1983)
[3] E Forest and R D Ruth, *Physica* **D43**, 105 (1990)
[4] Feng Kang, *Proc. 1984 Beijing Symposium on Differential Geometry and Differential Equations* edited by Feng Kang (Science Press, Beijing, 1985), p. 42

[5] F Neri, *Lie algebras and canonical integration*, Department of Physics Technical Report, University of Maryland (1988)
[6] P J Channell and C Scovel, *Nonlinearity* **3**, 231 (1990)
[7] H Yoshida, *Phys. Lett.* **A150**, 262 (1990)
[8] J Irwin, SSC Report, No. 228 (1989)
[9] G Rangarajan, Ph.D. thesis, University of Maryland (1990)
[10] G Rangarajan, *J. Math. Phys.* **37**, 4514 (1996)
[11] G Rangarajan *et al*, *Part. Accel.* **28**, 119 (1990)
[12] A J Dragt and D T Abell, *Int. J. Mod. Phys.* **2B**, 1019 (1993)
[13] D T Abell, Ph.D. thesis, University of Maryland, 1996
[14] F Lasagni, *ZAMP* **39**, 952 (1988)
[15] Y B Suris, *U.S.S.R. Comp. Maths. Math. Phys.* **29**, 149 (1989)
[16] A Iserles, *Bull. Greek Math. Soc.* **32**, 3 (1991)
[17] J M Sanz-Serna, *Acta Numerica* **1**, 243 (1992)
[18] D Okubnor and R D Skeel, *SIAM J. Numer. Anal.* **29**, 521 (1992)
[19] J M Sanz-Serna and M P Calvo, *Numerical Hamiltonian Problems* (Chapman and Hall, London, 1994)
[20] G P Ramaswami, Ph.D. thesis, Cambridge University, 1996
[21] H Goldstein, *Classical Mechanics* second edition (Addison-Wesley, New York, 1980)
[22] A J Dragt, in *Physics of high energy particle accelerators* edited by R A Carrigan, F R Huson and M Month, AIP Conference Proceedings No. 87 (American Institute of Physics, New York, 1982), p. 147
A J Dragt *et al*, *Ann. Rev. Nucl. Part. Sci.* **38**, 455 (1988)
[23] J F Cornwell, *Group Theory in Physics* (Academic Press, New York, 1984)
[24] A J Dragt and J M Finn, *J. Math. Phys.* **17**, 2215 (1976)
[25] P J Channell, Explicit integration of kick Hamiltonians in three degrees of freedom, AT-6:ATN-86-6, Los Alamos National Laboratory (1986)
[26] I Gjaja, *Part. Accel.* **43**, 133 (1994)

physics pp. 143–161

Painlevé analysis and integrability aspects of some nonlinear partial differential equations

K PORSEZIAN
Institut für Theoretische Physik, Universität Hannover, 30167 Hannover, Germany and
Centre for Laser Technology, Department of Physics, Anna University, Madras 600025, India

Abstract. A brief review of the Painlevé singularity structure analysis of some autonomous and nonautonomous nonlinear partial differential equations is discussed. We point out how the Painlevé analysis of solutions of these equations systematically provides the integrability properties of the equation. The Lax pair, Bäcklund transformation and bilinear forms are constructed from the analysis.

Keywords. Painlevé analysis; integrability; solitons; Lax pair.

PACS Nos 02.30; 42.65; 43.25; 47.20; 52.35

1. Introduction and historical background

The subject of completely integrable models is fascinating. Decades of research in this area has led to several new physical and mathematical developments which are quite beautiful and which unify various aspects of physical problems that appear to be desperate. Mathematically, we encounter new concepts such as complete integrability, Lax pair, solitons and so on and several new methods have been developed. Among them the Painlevé singularity structure analysis is one of the systematic and powerful method in nonlinear science to identify the integrability case(s) and the complete integrability properties of nonlinear systems. From the physics point of view, the models discussed in this article describe physical phenomena in such diverse areas as nonlinear optics, hydrodynamics, condensed matter, plasma physics, relativistic theory and so on. In this review article we will discuss the Painlevé analysis of various nonlinear models which are completely integrable. These models describe systems of nonlinear differential equations which can be solved exactly. Most of these models would be continuum models (both one and higher dimensions as well as autonomous and nonautonomous)although we will also investigate the Painlevé analysis of the Toda lattice and some nonintegrable systems. Before analysing the singularity structure analysis of these equations, first we will briefly discuss the historical background of Painlevé analysis.

Determining whether or not a given system of nonlinear ordinary or partial differential equations is integrable raises many fundamental issues. In particular there is the issue of what is meant by 'integrability' and the issue of how that integrability can best be determined without having to resort to a complete solution of the problem. For Hamiltonian systems the notion of integrability is well defined, i.e. the existence of as many involutive first integrals as there are degrees of freedom. For non-Hamiltonian systems things are less clearcut. Clearly the existence of integrals can lead to a reduction of the order of the system and hence to a solution in terms of an 'integration by quadratures'. However, this is clearly not the whole story since there are simple, apparently 'integrable', equations such as the Painlevé transcendents which do not have algebraic integrals and for which an integration by quadratures is not possible.It now seems that it is possible to identify many different classes of integrable systems on the basis of their analytic structure, i.e. the types of singularities exhibited by their solutions in the complex domain. The techniques for doing this can be applied to both the ordinary and partial differential equations and furthermore can be extended to provide, in many cases, explicit solutions to the systems in question. It also turns out that the complex domain of nonintegrable systems also contains much valuable information and overall it would seem that the study of analytic structure can provide a wide-ranging and unified treatment for a large class of nonlinear problems. Historically, the first application of these ideas is due to Kovalevskaya [1] in her famous work on the rigid body problem. This concerned the then popular problem of trying to find a general solution to the Euler–Poisson equations which describe the motion of a top spinning about a fixed point. Her approach was mainly to determine the conditions under which the only movable singularities, i.e. those singularities whose positions are initial-condition-dependent, exhibited by the solutions to the equations of motion, in the complex plane, are ordinary poles. She found that this only occured for four special combinations of the adjustable system parameters (the moments of inertia and the position of the centre of gravity). She was, able to identify the known integrable cases and one new one. Following Kovalevskaya came the extensive work of Painlevé [2] and co-workers to determine the categories of second order equations whose only movable singularities are ordinary poles. Working with the general class of equations

$$\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = F\left[\frac{\mathrm{d}y}{\mathrm{d}x}, y, x\right], \tag{1}$$

where F is analytic in x and rational in y and $\mathrm{d}y/\mathrm{d}x$, Painlevé found that there were 50 types which had the desired analytical property. Forty-four of these equations are solvable in terms of known functions. The reamining six equations, referred to as the Painlevé transcendents [3], have transcendental meromorphic solutions for which convergent expansions are explicitly known. Despite their mathematical interest it appeared, for many years, that the Painlevé transcendents were devoid of physical content. However, in the last two decades or so they have started to reappear in various important physical contexts such as the work of Wu *et al* [4] which showed that two-point correlation function of the Ising model satisfied the third Painlevé transcendent. In the last few years there is a growing very large number of literature on the Painlevé property as a means of identifying integrable cases of nonlinear ordinary, differential difference and partial differential equations. For example, this approach has been used successfully to predict integrable cases of a variety of systems such as the Lorentz equations [5], the Henon–Heiles system [6], the three particle Toda lattice [7], many systems in soliton theory [8–14], etc. It has also been observed that there are integrable systems with movable rational branch points. This has led to the concept of a so-called 'weak Painlevé property' [8].

For partial differential equations (PDE) 'complete integrability' is usually taken to mean the existence of an infinite number of conservation laws. As is well known such systems can sometimes exhibit N-soliton solutions and be solved by the inverse scattering transform(IST) method [9]. As with ODEs there is also important question of finding direct tests of complete integrability. Constructive approaches include using the method of differential forms, the use of Hirota's direct method [10] for constructing N-soliton solutions and certain criteria based on the properties of ODEs obtained by various reductions of the PDEs in question. This latter approach has been developed by Ablowitz, Ramani and Segur [11] and is based on the idea that all the ODEs obtained by similarity and travelling wave reductions from the PDE, should possess the Painlevé property. This conjecture has proved to be quite valuable, although there are clearly certain drawbacks such as the difficulty of identifying all possible reductions of the PDEs to ODEs and inability of the conjecture to provide further information about actual solutions to the equation in hand. In order to overcome these difficulties Weiss, Tabor and Carnevale [12] who have formulated a Painlevé (P-) type test that can be applied directly to PDEs without any need for reductions. This approach does seem to provide a valuable first test and, in addition, seems to be capable of yielding other important informations such as Bäcklund transformations, Lax pair, rational solutions etc. The major difference between the P-analysis of ODEs and PDEs is that the singularities of the latter are in general not isolated but lie on some analytic manifold, the singular manifold, determined by conditions of the form

$$\phi(x_1, x_2, \ldots, x_n) = 0, \tag{2}$$

where ϕ is analytic in the neighbourhood of that manifold. To show that a given equation, say

$$q_t = K(q), q = q(x,t), \tag{3}$$

possesses the generalized P-property requires a demonstration that the solutions may be expanded locally in a Laurent-like series about the singular manifold in the complex hyperspace, i.e. an expansion of the form

$$q(x,t) = \phi^{-\alpha}(x,t) \sum_{j=0}^{\infty} q_j(x,t)\phi^j(x,t), \tag{4}$$

where α is some(integer) leading order and the $q_j(x,t)$ a set of expansion coefficients analytic in the neighbourhood of the singular manifold $\phi(x,t) = 0$. Such an expansion will have certain values of j, termed 'resonances', at which the corresponding q_j should be arbitrary. One resonance always occurs at $j = -1$ and corresponds to the arbitrariness of ϕ itself and the positions of the other resonances are determined by the nonlinearities in $K(q)$.

By Cauchy–Kovalevskaya's theorem such an expansion of the general solution must have sufficient number of arbitrary functions equal to that of the order of the PDEs. Furthermore, depending on the form of $K(q)$, the equation may also have different leading orders, each one of which will have its own resonance structure. We usually call the leading order corresponding to the general solution the 'principle branch' and the others (with less arbitrary functions) as 'lower branches'. If, at a resonance (of any branch) the associated q_j fails to be arbitrary, terms of the form $\phi^j \ln \phi$ must be included in the expansion. This makes the solution multi-valued about the singular manifold and hence the P-property is lost.There are essentially four steps involved in the P-analysis of PDEs: (i) Determination of the leading-order behaviours; (ii) Identification of the powers at which arbitrary functions can enter into the Laurent series called resonances; (iii) Verifying that at each resonance values sufficient number of arbitrary functions exist without the introduction of movable manifolds; (iv) Establishing connections with the Lax pair, Bäcklund transformation (BT), bilinear form and other integrability properties.

2. Painlevé analysis of autonomous nonlinear partial differential equations

2.1 *Painlevé analysis of modified Korteweg de Vries equation*

A simple illustration is provided by the MKDV equation [12]

$$q_t - 6q^2 q_x + q_{xxx} = 0. \tag{5}$$

The leading order is easily determined to be -1 with $q_0 = \phi_x$. For finding the resonances, we substitute

$$q = \sum_{j=0} q_j \phi^{j-1}, \tag{6}$$

into (5) and equating the coefficients of ϕ^{j-4} , we get the resonance values in the form

$$j = -1, 3, 4. \tag{7}$$

The resonance $j = -1$ represents the arbitrariness of the singular manifold $\phi(x,t) = 0$. From the other powers of ϕ, one finds that

$$j = 0; \quad q_0 = \phi_x; \tag{8}$$

$$j = 1; \quad q_1 = -\frac{\phi_{xx}}{2\phi_x}; \tag{9}$$

$$j = 2; \quad \phi_t - 6q_2\phi_x^2 - \frac{3\phi_{xx}^2}{2\phi_x} + \phi_{xxx} = 0; \tag{10}$$

$j = 3; q_3$ arbitrary if

$$\frac{\partial}{\partial x}\left(\phi_t - 6q_2\phi_x^2 - \frac{3\phi_{xx}^2}{2\phi_x} + \phi_{xxx}\right); \tag{11}$$

$j = 4; q_4$ arbitrary. Clearly the 'compatibility condition' at $j = 3$ is always satisfied by virtue of the relations found at $j = 2$. Thus the expansion (6) is valid and we can say that the MKDV equation passes the P-test. To simplify the computations one can use a modification of this method proposed by Kruskal [13]. This consists of expanding ϕ about the singular manifold in the form $\phi(x,t) = x + \psi'(t)$, such that $\phi(\psi'(x,t)) = 0$, and setting $q_j = q_j(t)$. For the MKDV equation this reduced approach yields $q_0 = 1, q_1 = 0, q_2 = \psi_t/6, q_3$ arbitrary, q_4 arbitrary. From the point of view of the 'test' alone, this modification is rather useful. However, to construct the integrability properties is easily shown to pass the Painlevé test [12–15]. The leading order is -2 and the resonances occur at $j = -1, 4, 6$. The truncated expansion then takes the form

$$q = \frac{\partial^2}{\partial x^2} \log \phi + q_2 \tag{13}$$

with

$$\phi_x \phi_t + 12 q_2 \phi_x^2 + 4 \phi_x \phi_{xxx} - 3 \phi_{xx}^2 = 0, \quad \phi_{xt} + 12 q_2 \phi_{xx} + \phi_{xxxx} = 0, \tag{14}$$

and

$$q_{2,t} + 12 q_2 q_{2,x} + q_{2,xxx} = 0. \tag{15}$$

The set of equations (14) and (15) constitute an auto BT for the KDV equation. It should be noted that the second logarithmic derivative in (13) is reminiscent of the transformation used in Hirota's method [10]. By making the substitution $\phi_x = \psi^2$ leads to

$$\psi_{xx} + (2q_2 + \lambda)\psi = 0, \quad \psi_t + (6q_2 + \lambda)\psi_x + \psi_{xxx} = 0, \tag{16}$$

which are precisely the Lax pair for the KDV equation [9]. One of the most enduring techniques in the study of integrable nonlinear PDEs is Hirota method for constructing N-soliton solutions [10]. The method proceeds without the knowledge of the IST and in some cases was used to construct N soliton solutions before the scattering transform had been found [10]. To get the bilinear transformation from the P- analysis, we have to assume that the constant term in the expansion is equal to zero. i.e. in the case of KDV, $q_2 = 0$. Then, we get

$$f f_{xxxx} - 4 f_x f_{xxx} + 3 f_{xx}^2 + f f_{xt} - f_x f_t = 0 \tag{17}$$

which can be written in terms of Hirota's bilinear operators as

$$(D_x^4 + D_x D_t) f \cdot f = 0, \tag{18}$$

where $D_x^n(f \cdot f) = (\partial_x - \partial_{x'})^n f(x) f(x') \mid x = x'$. As is well known f can be developed in a series expansion which self truncates at each order yielding $1, 2, 3, \ldots$, N-soliton solutions. It is generally believed that if one can construct N-soliton solutions for $N \geq 3$, then they will exist for all N and the system can be deemed integrable. It should be noted that many equations can be reduced to bilinear form and one can often construct one and two soliton solutions for them by this method. However, self-truncation for all N only seems to occur only for completely integrable systems.

2.2 *Painlevé analysis of Pohlmeyer–Lund–Regge equation*

Lund *et al* [16] have shown that the dynamics of relativistic vortices (equivalently, strings) interacting through a scalar field has led to a set of two coupled, Lorentz-invariant, nonlinear equations in two independent variables. The equations are in the form

$$\Phi_{xt} - \sin\Phi\cos\Phi - \frac{sin\Phi}{\cos^3\Phi}\Psi_x\Psi_t = 0, \tag{19}$$

$$(\Psi_t\tan^2\Phi)_x + (\Psi_x\tan^2\Psi)_t = 0. \tag{20}$$

Eqs. (19–20) are conditions for embedding a 2-dimensional surface in a three dimensional sphere which itself embedded in a four dimensional Euclidean space. Eqs. (19) and (20) were also derived by Pohlmeyer [17] through a study of the nonlinear σ- models of field theory.

For convenience, we use the transformation given by Getmanov [18] in the form

$$q = \sin\Phi\exp(i\Psi). \tag{21}$$

Under this transformation, eqs. (19) and (20) are transformed into

$$q_{xt} + \frac{q_x q_t q^*}{1-|q|^2} - q(1-|q|^2) = 0. \tag{22}$$

To apply P-analysis, we define $q = a$ and $q^* = b$. Eq. (22) in terms of these change of variables takes the form [19]

$$a_{xt} + \frac{a_x a_t b}{1-ab} - a(1-ab) = 0, \quad b_{xt} + \frac{b_x b_t a}{1-ab} - b(1-ab) = 0. \tag{23}$$

Leading order: $\alpha = \beta = -1$ with $a_0 b_0 = -\psi_t$. Resonances: $j = -1, 0, 1, 2$. Collecting the coefficients of ϕ^{-4} and ϕ^{-4}, we obtain

$$a_0 b_1 + a_1 b_0 = 0. \tag{24}$$

From eq. (24) it is clear that either a_1 or b_1 is arbitrary. In a similar way proceeding further by collecting the coefficients of ϕ^{-3} and ϕ^{-3}, we get

$$(a_0 b_2 - b_0 a_2)\psi_t = a_{0t} b_1 + b_0 a_{1t}, (b_0 a_2 - a_0 b_2)\psi_t = b_{0t} a_1 + a_0 b_{1t}. \tag{25}$$

Using (24), it is easy to show that the functions a_2 or b_2 is arbitrary. Thus the general solution $(a(x,t), b(x,t))$ of eq. (22) admits the sufficient number of arbitrary functions without the introduction of any movable critical manifold, thus satisfying the P- property and hence the system is expected to be integrable. As eq. (22) satisfies the required condition to be integrable, we now proceed to obtain the associated integrability properties of the system. Now to establish the integrability properties, the series representation is truncated at the constant level term $(a_j = b_j = 0, j > 1)$ as

$$a = \frac{a_0}{\phi} + a_1, \quad b = \frac{b_0}{\phi} + b_1, \tag{26}$$

where a_0 and b_0 satisfy eqs (23). Equation (26) can also be treated as an auto BT of eq. (22). Then to find the bilinear form, we put $a_1 = b_1 = 0$ and then by defining

$$a = \frac{a_0}{\phi} = \frac{G}{F}, \quad b = \frac{b_0}{\phi} = \frac{G^*}{F} \tag{27}$$

solution to it and also conjectured N soliton solutions. To construct the bilinear form from (28), we define

$$F^2 = f^*f + g^*g. \tag{29}$$

Under this transformation eqs (28) are transformed into the bilinear equations

$$D_x(f^* \cdot f + g^* \cdot g) = 0, (D_xD_t - 1)f \cdot g^* = 0, D_xD_t(f^* \cdot f - g^* \cdot g) + 2g^*g = 0. \tag{30}$$

Once the bilinear forms are known, then one can generate the soliton solutions by expanding the dependent variables in terms of power series.

2.3 *Some nontrivial equations and Painlevé condition(s)*

It would seem that all the known integrable PDEs pass the Painlevé test. Some nontrivial examples include

I. Hirota–Satsuma equation [14]

$$q_t - \lambda(6qq_x + q_{xxx}) = -6pp_x, \quad p_t + 3qp_x + p_{xxx} = 0 \tag{31}$$

which is found to pass the P-test only when $\lambda = \frac{1}{2}$.

II. Unidirectional Zakharov equation [20]

$$iq_t + q_{xx} = qp, \quad p_t + p_x = (| q |^2)_x. \tag{32}$$

III. Nonlinear Schrödinger-Maxwell Bloch equations [21]

$$q_x = ikq_{tt} - ig \mid q \mid^2 q + 2\alpha_2\langle p\rangle, \quad p_t = i\omega p + fq\eta, \quad \eta_t = 2f(qp^* + q^*p). \tag{33}$$

Painlevé condition: $-2f^2k = g$.

IV. Higher order nonlinear Schrödinger equation [22]

$$iq_z + \alpha_1 q_{tt} + \alpha_2 \mid q \mid^2 q - i\varepsilon[\alpha_3 q_{ttt} + \alpha_4(q^2q^*)_t + \alpha_5 q(qq^*)_t] = 0. \tag{34}$$

Painlevé conditions: $\alpha_2 = 2\alpha_1, \alpha_3 : \alpha_4 : \alpha_5 = 1 : 6 : -3$.

2.4 *Non-Painlevé property of the integrable extended nonlinear Schrödinger equation*

Recently Liu *et al* [23] have shown that the extended nonlinear Schrödinger equation is a completely integrable system and constructed the N-soliton solutions through Hirota's bilinear method. The equation is of the form

$$iq_x + \frac{1}{2}q_{tt} + \frac{\mid q \mid^2 q}{2} - i\delta q_{ttt} + 2i\alpha_1 \mid q \mid^2 q_t + i\alpha_1 \mid q \mid^2 q_t^* = 0. \tag{35}$$

$$ia_t + \tfrac{1}{2}a_{xx} + \tfrac{1}{2}a^2 b + i[-\delta a_{xxx} + 2\alpha_1 a b a_x + \alpha_1 a^2 b_x] = 0, \tag{36}$$

$$-\,ib_t + \tfrac{1}{2}b_{xx} + \tfrac{1}{2}ab^2 - i[-\delta b_{xxx} + 2\alpha_1 a b b_x + \alpha_1 b^2 a_x] = 0. \tag{37}$$

Leading Order: $\alpha = \beta = -1$; $a_0 b_0 = \frac{2\delta}{\alpha_1}$. Resonances: $j = -1, 0, 3, 3, 3, 4$. The resonance '-1' corresponds as usual to the arbitrariness of the singular manifold and '0' corresponds to the fact that either a_0 or b_0 is arbitrary as seen from the leading order coefficient. As the system admits more resonances at $j = 3, 3, 3$ than the required arbitrary functions, we conclude that eq. (35) fails to satisfy the Painlevé test.

In the cases considered so far reveals that if the P-property is satisfied, then the equation is completely integrable. Thus it is conjectured that this property indicates the solvability of a field equation. On the other hand we cannot conclude, in general, that a PDE which is completely integrable has the P-property. Examples are the Harry Dym equation $q_t = q^3 q_{xxx}$ and the nonlinear diffusion equation $q_t = (q^{-2} q_x)_x$ [14]. Both equations are integrable. The first can be solved by the IST and the second can be linearized to the linear diffusion equation.

2.5 *Painlevé analysis of nonintegrable equations*

Painlevé analysis is also very useful to understand the dynamics of the nonintegrable evolution equations. For example, for the real Newell–Whitehead equation [25]

$$q_t = q_{xx} + q - 2q^3 \tag{38}$$

the leading order is -1 and the resonances are -1 and 4. For simplicity, we define new variables $w = \phi_t/\phi_x$, $v = \phi_{xx}/\phi_x$ and from the arbitrary analysis, we find

$$j = 0 : q_0 = \phi_x, \tag{39}$$

$$j = 1 : q_1 = \tfrac{1}{2}(w - 3v), \tag{40}$$

$$j = 2 : q_2 = \frac{1}{6\phi_x}\left(v_x - w_x + 1 - \tfrac{1}{6}w^2 - \tfrac{1}{2}v^2\right), \tag{41}$$

$$\begin{aligned} j = 3 : q_3 = \frac{1}{24\phi_x^2}\Big(&-w_t + 2w_{xx} + \tfrac{7}{3}ww_x - 2w + \tfrac{4}{9}w^3 \\ &- v_{xx} + 3vv_x + 2v - v^3 - 2vw_x - \tfrac{1}{3}w^2 v\Big), \end{aligned} \tag{42}$$

$$j = 4 : 0(q_4) = \frac{1}{3\phi_x^3}\left(w_t - ww_x + \tfrac{3}{2}w - \tfrac{1}{3}w^3\right). \tag{43}$$

From eq. (43), it is clear that the Laurent expansion fails the WTC test since q_4 is found not to be arbitrary. We therefore have the following "consistency condition" (assuming $\phi_x \neq 0$):

$$w_t - ww_x + \tfrac{3}{2}w - \tfrac{1}{3}w^3 = 0. \tag{44}$$

There are two possible ways to proceed: (i) either, introduce logarithmic psi-series, i.e.

$$q = \sum_j \sum_k q_{jk} \phi^{j-1} (\phi^4 \ln \phi)^k. \tag{45}$$

(ii) or, force ϕ to satisfy the compatibility condition at $j = 4$. This takes the form of a nonlinear PDE which can be solved exactly.This approach gives a type of conditional P-property and can yield special type of solutions meromorphic in the singular manifold which is now specified as opposed to being arbitrary. In view of the above facts, we can conclude that the real Newell–Whitehead equation has the "conditional Painlevé property". It is also interesting to note that some of the special techniques developed for integrable systems are also shown to be applicable to nonintegrable equations. Let us consider the second example from one parameter continuous (Lie) group of transformations point of view [26]. For this, we consider the nonintegrable KdV equation in the form

$$q_t + q^n q_x + q_{xxx} = 0, \quad n > 2. \tag{46}$$

Using the invariant variables

$$\zeta = \frac{(x + \frac{3\beta}{\alpha})}{(\alpha + \beta)^{1/3}} \tag{47}$$

and

$$q = \frac{f(\zeta)}{(\alpha t + \delta)^{2/3n}} \tag{48}$$

in eq. (46), we obtain the invariant equation in the form

$$f''' + f^n f' - \left(\frac{2f}{n} + \zeta f'\right) = 0 \tag{49}$$

Inserting

$$f \sim \alpha(\zeta - \zeta_0)^p \tag{50}$$

in eq. (49), we find that there is only one possibility

$$p = \frac{-2}{n}, \alpha = \left(\frac{-(2+n)(2+2n)}{n^2}\right)^{1/n}. \tag{51}$$

Since $n > 2$, eq. (49) will have a movable branch point of order $-2/n$ provided (50) is asymptotic near ζ_0. To see this asymptotic nature, we define

$$f = v^{-2/n} \tag{52}$$

The equation for v is

$$-\frac{2}{n}v^2 v''' + \frac{6}{n}\left(\frac{2}{n}+1\right)vv'v'' - \frac{2}{n}\left(\frac{2}{n}+1\right)\left(\frac{2}{n}+2\right)v'^3 - \frac{2}{n}v' - \frac{2}{n}v^3 + \frac{2\zeta}{n}v^2 v' = 0 \tag{53}$$

There is a regular solution of (53), that is, regular at ζ_0, if $v(\zeta_0)=0, v'(\zeta_0)+(2+n)(2+2n)/n^2\ (v'(\zeta_0))^3=0$. $v''(\zeta_0)$ is a finite quantity,and $v'''(\zeta_0)$ is finite. Then $v(\zeta)$ is analytic at ζ_0 and so (50) is asymptotic near ζ_0. Thus eq. (49) is not of Painlevé type. For more information about the P-analysis of nonintegrable systems, we refer the reader to refs. [25, 26].

3. Painlevé property for differential-difference equations

In this section, we will briefly discuss the Painlevé analysis of the discrete systems. The Toda lattice [27]

$$Q_{n,tt} = e^{Q_{n-1}-Q_n} - e^{Q_n-Q_{n+1}} \tag{54}$$

is probably the best known member of a class of differential-difference equations that are integrable. To establish the P-property [27], we define the variables $q_n = Q_{n,t}$, $p_n = e^{Q_n-Q_{n+1}}$, i.e.

$$p_{n,t} = p_n(q_n - q_{n+1}), \quad q_{n,t} = p_{n-1} - p_n. \tag{55}$$

A solution of these equations was found by Toda on making the substitution

$$p_n = \frac{\partial^2}{\partial t^2} \log f_n \tag{56}$$

which immediately leads to

$$q_n = \frac{\partial}{\partial t} \log\left(\frac{f_{n-1}}{f_n}\right) \tag{57}$$

and

$$\left[\frac{f_{n+1}f_{n-1}}{f_n^2} - 1\right] = \frac{\partial^2}{\partial t^2} \log f_n. \tag{58}$$

The one soliton solution corresponds to the choice $f_n \cosh(\alpha n - (\sinh \alpha)t)$. These results suggests a truncated 'Painlevé type' expansion of the form

$$p_n = \frac{p_n^{(0)}}{\phi_n^2} + \frac{p_n^{(1)}}{\phi_n} + p_n^{(2)}, \quad q_n = \frac{q_{n-1}^{(0)}}{\phi_{n-1}} - \frac{q_n^{(0)}}{\phi_n} + q_n^{(1)} \tag{59}$$

from which one can easily determine that $q_n^{(0)} = \phi_{n,t}, p_n^{(0)} = -\phi_{n,t}^2$ and $p_n^{(1)} = \phi_{n,tt}$, i.e.

$$p_n = \frac{\partial^2}{\partial t^2} \log \phi_n + p_n^{(2)}, \quad q_n = \frac{\partial}{\partial t} \log\left(\frac{\phi_{n-1}}{\phi_n}\right) + q_n^{(1)} \tag{60}$$

and that

$$p_n^{(2)}\left[\frac{\phi_{n+1}\phi_{n-1}}{\phi_n^2} - 1\right] = \frac{\partial^2}{\partial t^2} \log \phi_n. \tag{61}$$

Thus, in a sense we are expanding each dependent variable p_n about its own 'singular manifold' function ϕ_n. The same approach for the two dimensional Toda lattice,

$$p_{n,x} = p_n(q_n - q_{n+1}), \quad q_{n,t} = p_{n-1} - p_n. \tag{62}$$

yields

$$p_n = \frac{\partial^2}{\partial t \partial x} \log \phi_n + p_n^{(2)} \tag{63}$$

which immediately leads to

$$q_n = \frac{\partial}{\partial x} \log\left(\frac{\phi_{n-1}}{\phi_n}\right) + q_n^{(1)} \tag{64}$$

and

$$p_n^{(2)}\left[\frac{\phi_{n+1}\phi_{n-1}}{\phi_n^2} - 1\right] = \frac{\partial^2}{\partial t \partial x} \log \phi_n. \tag{65}$$

From the relations (63) and (64) one can easily determine Lax pairs for the corresponding systems. For example, in the case of the 2-D lattice the scattering scheme is found to be

$$\phi_{n,x} = \alpha_{n+1}\phi_n + \beta_n\phi_{n+1}, \quad \phi_{n,t} = \gamma_n\phi_{n-1} + \delta_{n-1}\phi_n \tag{66}$$

and on substituting into (65), combined with cross differentiation of (66), deduce that $\beta_n = \lambda$ (spectral parameter), $\alpha_n = -q_n^{(1)}$ and $\gamma_n = \frac{p_n^{(2)}}{\lambda}$. This type of P-analysis also yields the BT's for the Toda lattices and reveals some amusing connections with the rational solutions of the second Painlevé transcendent [27].

4. Painlevé analysis of nonautonomous nonlinear partial differential equations

The problem of nonlinear wave propagation in dispersive and inhomogeneous media has been of general interest and has had wide range of applications, e.g. radio waves in the ionosphere, waves in the ocean,nonlinear optics, magnetic systems, etc, [31–35]. For the past few years, several inhomogeneous nonlinear PDEs have been studied from the Painlevé property and soliton point of view. In a series of papers, we have systematically investigated the Painlevé analysis of many of these equations and explained the construction of soliton solutions [35–40]. In this section, we will briefly discuss the P-analysis of some nonautonomous equations. As pointed out in section I, to identify the Painlevé condition(s) of these systems, we have to use the general manifold.

4.1 *Painlevé analysis of inhomogeneous spherically symmetric nonlinear Schrödinger equation*

First, we will discuss the P-analysis of the inhomogeneous spherically symmetric nonlinear Schrödinger equation of the form [35]

$$\begin{aligned} iq_t + (fq)_{rr} + \left(f\frac{n-1}{r}q\right)_r + 2f \mid q \mid^2 q \\ + \left[2\int_0^r f_{r'} \mid q \mid^2 \mathrm{d}r' + 4(n-1)\int_0^r \frac{f}{r'} \mid q \mid^2 \mathrm{d}r'\right] q = 0. \end{aligned} \tag{67}$$

The above equation can be derived from the inhomogeneous spherically symmetric continuum Heisenberg ferromagnet in arbitrary ($n-$) dimensions through moving helical space curve [35]. Equation (67) contains the following integrable models:

(i) $n=0, f=$ constant: nonlinear Schrödinger equation

$$iq_t + q_{xx} + 2\mid q\mid^2 q = 0(-\infty < x < \infty). \tag{68}$$

(ii) $n=1, f=\alpha x+\beta$: deformed nonlinear Schrödinger equation

$$iq_t + (\alpha x+\beta)(q_{xx} + 2\mid q\mid^2 q) + 2\alpha\left(q_x + q\int_0^x \mid q\mid^2 \mathrm{d}x'\right) = 0. \tag{69}$$

(iii) $n=2, f=$ constant : deformed radial nonlinear Schrödinger equation

$$iq_t + q_{rr} + \frac{1}{r}q_r - \frac{1}{r^2}q + 2\mid q\mid^2 q + 4q\int_0^r \frac{1}{r}\mid q\mid^2 dr' = 0. \tag{70}$$

After removing the integral through new variable R and by denoting $q=a$ and $q^*=b$, eq. (67) can be written as

$$ia_t + f\left(a_{rr} + \frac{n-1}{r}a_r - \frac{n-1}{r^2}a\right) + 2Ra + 2f_r a_r + \left[f_{rr} + \frac{n-1}{r}f_r\right]a = 0, \tag{71}$$

$$-ib_t + f\left(b_{rr} + \frac{n-1}{r}b_r - \frac{n-1}{r^2}b\right) + 2Rb + 2f_r b_r + \left[f_{rr} + \frac{n-1}{r}f_r\right]b = 0, \tag{72}$$

$$R_r - 2f_r ab - f\left[ab_r + a_r b + 2\frac{n-1}{r}ab\right] = 0. \tag{73}$$

Leading order: $\alpha=\beta=-1$, $\gamma=-2$, $a_0b_0=\phi_r^2$, $R_0=-f\phi_r^2$. Resonances: $j=-1, 0, 2, 3, 4$. From the detailed arbitrary analysis we find that eqs. (71–73) are free from movable critical manifolds only when f is of the form:

$$f = \varepsilon_1 r^{-2(n-1)} + \varepsilon_2 r^{-(n-2)}, \tag{74}$$

where ε_1 and ε_2 are integration constants.Hence, the inhomogeneous spherically symmetric nonlinear Schrödinger equation is expected to be integrable in arbitrary ($n-$) dimensions only when the inhomogeneity is of the form (74). Working with the truncated expansions

$$a = \frac{a_0}{\phi} + a_1, \quad b = \frac{b_0}{\phi} + b_1, \quad R = \frac{R_0}{\phi^2} + \frac{R_1}{\phi} + R_2, \tag{75}$$

the following overdetermined systems of equations are obtained: $O(\phi^{-3}, \phi^{-3}, \phi^{-3})$:

$$a_0b_0 = \phi_r^2, \quad R_0 = -f\phi_r^2. \tag{76}$$

$O(\phi^{-2}, \phi^{-2}, \phi^{-2})$:

$$-ia_0\phi_t - \left(2f_r + \frac{n-1}{r}f\right)a_0\phi_r - f(2a_{0r}\phi_r + a_0\phi_{rr}) + 2(a_0R_1 + R_0a_1) = 0, \tag{77}$$

$$ib_0\phi_t - \left(2f_r + \frac{n-1}{r}f\right)b_0\phi_r - f(2b_{0r}\phi_r + b_0\phi_{rr}) + 2(b_0R_1 + R_0b_1) = 0, \tag{78}$$

$$R_{0r} - R_1\phi_r - 2f_ra_0b_0 - f[(a_0b_0)_r - (a_0b_1 + a_1b_0)\phi_r] - 2\,\frac{n-1}{r}fa_0\,b_0 = 0. \tag{79}$$

$O(\phi^{-1}, \phi^{-1}, \phi^{-1})$:

$$ia_{0t} + (fa)_{rr} + \left(f\,\frac{n-1}{r}a\right)_r + 2(a_0R_2 + a_1R_1) = 0, \tag{80}$$

$$-ib_{0t} + (f\,b)_{rr} + \left(f\,\frac{n-1}{r}b\right)_r + 2(b_0R_2 + b_1R_1) = 0, \tag{81}$$

$$R_{1r} - 2f_r(a_0b_1 + a_1b_0) + f(a_0b_1 + a_1b_0)_r - 2\,\frac{n-1}{r}f(a_0b_1 + a_1 + b_0) = 0. \tag{82}$$

By noting that $(a_0\,b_{0t} + a_{0t}\,b_0) = -2\phi_r\phi_{rt}$ and using the expression for ϕ_t from eqs. (77–78) one may, after some simplifications, obtain

$$\left[\frac{a_{0r} + 2a_1\phi_r}{r^{n-1}a_0}\right]_r = \left[\frac{b_{0r} - 2b_1\phi_r}{r^{n-1}b_0}\right]_r. \tag{83}$$

Treating the constant of integration as the spectral parameter λ and identifying $a_0 = i\psi_1^2$, $b_0 = i\psi_2^2$, $a_1 = q$, $b_1 = q^*$, and $\phi_r = -i\psi_1\psi_2$, gives the scattering problem

$$\psi_{1r} = i\lambda r^{n-1}\psi_1 + q\psi_2, \quad \psi_{2r} = -q^*\psi_1 - i\lambda r^{n-1}\psi_2. \tag{84}$$

The time dependent part of the problem may be obtained directly from equations (80–82) by repeated use of (83) and (84), giving

$$\psi_{1t} = (iR - 2i\lambda^2 f\,r^{2(n-1)})\psi_1 + \left(-2\lambda f\,qr^{n-1} + i(fq)_r + i\frac{n-1}{r}fq\right)\psi_2, \tag{85}$$

$$\psi_{2t} = \left(2\lambda fq^*r^{n-1} + i(fq^*)_r + i\frac{n-1}{r}fq^*\right)\psi_1 + (-iR + 2i\lambda^2 fr^{2(n-1)})\psi_2. \tag{86}$$

Here f is given in eq. (74) and the consistency of the linear eigenvalue problem (84) and (85–86) is indeed the evolution equation (67) (only for the choice of f given in eq. (74)), provided λ evolves as

$$\lambda_t = (\varrho + i\xi)_t = 2n\varepsilon_2\lambda^2. \tag{87}$$

4.2 *Painlevé analysis of the generalized x-dependent MKDV equation*

We consider the GMKDV equation in the form [31]

$$\begin{aligned} q_t + \mu_1 q &+ (\nu_3 + \mu_3 x)(q_{xxx} - 6q^2 q_x) \\ &+ (\mu_1 x + \nu_1)q_x + \mu_3\left(3q_{xx} + 4q^3 - 2q_x \int_x^{\infty} \mathrm{d}x' q^2\right) = 0 \end{aligned} \tag{88}$$

In this equation ν_1, ν_3, μ_1 and μ_3 are real constants and q is a real. For $\mu_1 = \mu_3 = 0$, eq. (88) reduces to the well known MKDV equation. If instead $\mu_3 \neq 0$, eq. (88) is an integro differential equation; it can be reduced to a pure differential equation by changing the dependent variable from q to $\int_x^{\infty} q^2(x', t)\mathrm{d}x'$ or differentiating with respect to x after dividing by q_x. However, we can remove the integral term by defining a new dependent variable $R = \int_x^{\infty} q^2(x', t)\mathrm{d}x'$. Now eq. (88) can be rewritten as a set of coupled equations in the form [40]

$$\begin{aligned} q_t + \mu_1 q &+ (\nu_3 + \mu_3 x)(q_{xxx} - 6R_x q_x) + (\mu_1 x + \nu_1)q_x \\ &+ \mu_3(3q_{xx} + 4qR_x - 2q_x R) = 0, \end{aligned} \tag{89}$$

$$R_x = q^2 \tag{90}$$

to within an arbitrary function of t (which occurs when (90) is integrated),which can always be removed by a simple time dependent gauge transformation of q. Leading order: $\alpha = \beta = -1$ with $q_0^2 = -\phi_x^2$, $R_0 = \phi_x$.

Resonances: $j = -1, 1, 3, 4$. From the truncated series, we get

$$q = \frac{q_0}{\phi} + q_1, \quad R = \frac{R_0}{\phi} + R_1. \tag{91}$$

With suitable transformation for q_0 and q_1 [15, 35, 40], after some simplifications, we obtain

$$\psi_{1x} = -i\lambda\psi_1 + q\psi_2, \quad \psi_{2x} = q\psi_1 + i\lambda\psi_2, \tag{92}$$

$$\psi_{1t} = A\psi_1 + B\psi_2, \quad \psi_{2t} = C\psi_1 + A\psi_2, \tag{93}$$

where $A = -4i\lambda^3(\mu_3 x + \nu_3) - 2i\lambda[\mu_3(x\int q^2 \mathrm{d}x')_x + \nu_3 q^2] + i(\nu_1 + \mu_1 x)\lambda$, $B = 4\lambda^2(\mu_3 x + \nu_3)q + 2i\lambda[\mu_3(xq)_x + \nu_3 q_x] + \mu_3[-(xq)_{xx} + 2q(x\int q^2 \mathrm{d}x')_x] + \nu_3[-q_{xx} + 2q^3] - (\nu_1 + \mu_1 x)q$, $C = 4\lambda^2(\mu_3 x + \nu_3)q - 2i\lambda[\mu_3(xq)_x + \nu_3 q_x] + \mu_3[-(xq)_{xx} + 2q(x\int q^2 \mathrm{d}x')_x] + \nu_3[-q_{xx} + 2q^3] - (\nu_1 + \mu_1 x)q$ with $\lambda_t = -\mu_1\lambda - 4\mu_3\lambda^3$.

4.3 *Painlevé analysis of inhomogeneous deformed Kaup system*

The deformed Kaup system is in the form [39]

$$q_t + \tfrac{1}{2}q^2 + xqq_x + x\eta_x = 0, \quad \eta_t + 2\eta q + x\eta_x q + x\eta q_x + 3q_{xx} + xq_{xxx} = 0 \tag{94}$$

Leading order: $\alpha = -1$, $\beta = -2$, $q_0 = \pm 2\phi_x$, $\eta_0 = -2\phi_x^2$. Resonances: $j = -1, 2, 3, 4$. Lax pair:

$$\psi_{xx} + U\psi = 0, \quad \psi_t = A\psi_x + B\psi \tag{95}$$

with

$$U = \lambda^2 + \frac{i\lambda q}{2} + \frac{\eta}{4} - \frac{q^2}{16}, \quad A = -2ix\lambda - \frac{xq}{2}, \quad B = -\frac{A_x}{2} \tag{96}$$

The compatibility leads to eq. (94) only when λ satisfies

$$\lambda_t = -2i\lambda^2. \tag{97}$$

4.4 *Painlevé analysis of parametrically driven Sine–Gordon equation*

We consider the generalized sine-Gordon equation [41]

$$q_{tt} - q_{xx} + f(t)\sin q = 0. \tag{98}$$

It is well known that the sine-Gordon equation $q_{tt} - q_{xx} + \sin q = 0$, which is completely integrable, does not pass the Painlevé test directly, but after the transformation $V = e^{iq}$ [13, 41]. Under this transformation, eq. (98) takes the form.

$$V V_{tt} - V_t^2 - V V_{xx} + V_x^2 + \tfrac{1}{2} f(t)(V^3 - V) = 0. \tag{99}$$

Leading order: $\alpha = -2$, $V_0 = 4(\phi_x^2 - \phi_t^2)/f$

Resonances: $j = -1, 2$. At the resonance $j = 2$ we obtain the ODE

$$f\frac{d^2 f}{dt^2} = \left(\frac{df}{dt}\right)^2. \tag{100}$$

As expected this ODE has the P-property. The general solution is given by

$$f(t) = C_1 e^{C_2 t}, \tag{101}$$

where C_1 and C_2 are integration constants.Similar type of equation with damping was also investigated in [42].

4.5 *Some inhomogeneous equations and Painlevé condition(s)*

I. Generalized KDV equation: [43]

$$q_t + f(t) q q_x + g(t) q_{xxx} = 0. \tag{102}$$

Painlevé condition: $g(t) = f(t)(a_0 \int^t \mathrm{d}s f(s) + b_0)$, a_0 and b_0 are arbitrary constants.

II. Generalized KDV equation [44]:

$$q_t + q q_x + q_{xxx} + a(x,t) q = b(x,t). \tag{103}$$

Painlevé conditions: $a_x = 0, a_t + 2a^2 - b_x = 0$.

III. Cylinderical KDV equation [45]

$$q_t + q q_x + q_{xxx} + \frac{aq}{t} = 0. \tag{104}$$

Painlevé condition: $a = \frac{1}{2}$.

IV. Generalized nonlinear Schrödinger equation [43]

$$iq_t + \alpha(t)q_{xx} + \beta(t) \mid q \mid^2 q = 0. \tag{105}$$

Painlevé condition: $\alpha(t) = \beta(t)(a_1 \int^t \mathrm{d}s f(s) + b_1)$,$a_1$ and b_1 are arbitrary constants.

V. Damped Nonlinear Schrödinger equation [46]

$$iq_t + q_{xx} - \mid q \mid^2 q = a(x,t)q + b(x,t). \tag{106}$$

Painlevé conditions: $a = x^2(\beta_t/2 - \beta^2) + x\alpha_1 + \alpha_0 + i\beta, b = 0, \alpha_0(t), \alpha_1(t)$ and $\beta(t)$ are arbitrary real analytic functions.

VI. Cylinderical Nonlinear Schrödinger equation [47]

$$iq_t + q_{xx} + \mid q \mid^2 q = \frac{q}{2t}. \tag{107}$$

VII. Inhomogeneous nonlinear Schrödinger equation [38]

$$iq_t + (f\,q)_{xx} + 2f \mid q \mid^2 q + \left[2\int_0^x f_{x'} \mid q \mid^2 \mathrm{d}x'\right] q = 0 \tag{108}$$

Painlevé condition: $f = \alpha x + \beta$.

VIII. Generalized inhomogeneous nonlinear Schrödinger equation [37]

$$iq_t + (\alpha x + \beta)(q_{xx} + 2 \mid q \mid^2 q) + 2\alpha\left(q_x + q\int_0^x \mid q \mid^2 \mathrm{d}x'\right) + i\mu_1(xq)_x + i\nu(q_{xxx} + 6 \mid q \mid^2 q_x) = 0 \tag{109}$$

IX. Radially symmetric nonlinear Schrödinger equation [38]:

$$iq_t + q_{rr} + \left(\frac{n-1}{r}q\right)_r + 2 \mid q \mid^2 q + 4(n-1)q\int_0^r r' \mid q \mid^2 \mathrm{d}r' = 0 \tag{110}$$

Painlevé conditions: $n = 1, 2$.

X. Inhomogeneous Nonlinear Schrödinger–Maxwell Bloch equations [36]

$$q_x = ikq_{tt} - ig \mid q \mid^2 q + \alpha_2(x,t)q + \langle p \rangle, \quad p_t = i\omega p + fq\eta, \quad \eta_t = 2f(qp^* + q^*p). \tag{111}$$

Painlevé conditions: $-2f^2k = g$, $\alpha_2 = 1/2(x + x_0)$.

5. Painlevé analysis of some higher dimensional equations

5.1 *Painlevé analysis of Kodomtsev–Petviashvili equation*

The KP equation [12, 14]

$$q_{tx} + q_x^2 + qq_{xx} + \sigma q_{xxxx} + q_{yy} = 0, \quad \sigma = \pm 1 \tag{112}$$

possesses the P-property. Leading order: -2 with $q_0 = -2\phi_x^2$ Resonances: $-1, 4, 5, 6$. Then equating the various powers of ϕ to zero we obtain sets of equations and one can easily verify that the functions q_4, q_5 and q_6 are arbitrary and so the P-property holds.

From the truncated expansion, the Lax pair is found to be in the form

$$\sigma\psi_y + \psi_{yy} + q\psi = 0,\ \psi_t + 4\psi_{xxx} + 6q\psi_x + 3\left(q_x - \sigma\int_{\infty}^{x} q_y \mathrm{d}x'\right) = 0 \qquad (113)$$

the compatibility of which is the KP eq. (112). The Painlevé analysis of similar type of equations is also discussed in [48].

5.2 *Painlevé analysis of cylinderical Kadomtsev–Petviashvili equation*

The cylinderical Kadomtsev–Petviashvili equation is of the form [49]

$$q_{xt} + qq_{xx} + q_x^2 + q_{xxxx} + a(t)q_x + b(t)q_{yy} = 0 \qquad (114)$$

Painlevé conditions: $da/dt + 2a^2 = 0$, $db/dt + 4ab = 0$. Solving the above conditions,we get $a(t) = 0$, $b(t) =$ constant, or $a(t) = 1/2(t - t_0)$, $b(t) = b_0/(t - t_0)^2$.

5.3 *Painlevé analysis of Davey–Stewartson equation*

The higher dimensional version of the nonlinear Schrödinger equation [50]

$$iq_t - \sigma_1 q_{xx} + q_{yy} = \sigma_2 \mid q \mid^2 + 2\sigma_1\sigma_2 Rq,\ \sigma_1 R_{xx} + R_{yy} + (\mid q \mid^2)_{xx} = 0 \qquad (115)$$

where $\sigma_i = \pm 1$, $i = 1, 2$. Leading order: $\alpha = \beta = -1$, $\gamma = -2$ with $(a_0^2 + b_0^2) = 2/\sigma_2$ $(\sigma_1\phi_x^2 + \phi_y^2)$, $R_0 = -(2/\sigma_2)\phi_x^2$. Resonances: $-1, 0, 2, 3, 3, 4$. From the truncation, the bilinear form is constructed as

$$(iD_t - \sigma_1 D_x^2 + D_y^2 - \sigma_2\lambda^2)g \cdot F = 0(\sigma_1 D_x^2 + D_y^2 - \sigma_2\lambda^2)F \cdot F = -\sigma_1 G \cdot G \qquad (116)$$

where λ = constant.

6. Summary and conclusions

In this short review, we have briefly discussed the Painlevé analysis and integrability properties of some autonomous and nonautonomous nonlinear partial differential equations. For a class of physically important equations we have demonstrated that this technique is one of the very useful and powerful method in nonlinear science to establish the integrability properties like Lax pair, Bäucklund transformation, bilinear form and so on. In this article, I have not included the Painlevé analysis of the coupled nonlinear Schrödinger equations. For coupled systems, this analysis is found to be cumbersome and the construction of Lax pair (from the truncation) is also an unsolved problem.

Acknowledgements

KP is grateful to Professor H J Mikeska for his very kind hospitality and constant encouragement. He expresses his thanks to DAAD for offering the fellowship, and to

DST, and CSIR, Govt. of India, for the financial support through major research projects.
This paper is dedicated to Professor M Lakshmanan on the eve of his 50th birthday.

References

[1] S Kovalevskaya, *Acta Math.* **12**, 177, (1889); **14**, 81(1889)
[2] P Painlevé, *Oeuvres*, Vol. I, CNRS, Paris (1973)
[3] E L Ince, *Ordinary differential equations* (Dover, New York, 1956)
[4] T T Wu, B M McCoy, C A Tracy and E Barouch, *Phys. Rev.* **B13**, 316 (1976)
[5] M Tabor and J Weiss, *Phys. Rev.* **A24**, 2157 (1981)
[6] Y F Chang, M Tabor and J Weiss, *J. Math. Phys.* **23**, 531 (1982)
[7] T Bountis, H Segur and F Vivaldi, *Phys. Rev.* **A25**, 1257 (1982)
[8] A Ramani, B Dorizzi and B Grammaticos, *Phys. Rev. Lett.* **49**, 1539 (1982)
[9] M J Alowitz, D J Kaup, A C Newell and H Segur, *Stud. Appl. Math. Phys.* **53**, 249 (1974)
[10] R Hirota, in *Solitons* edited by R K Bullough and P J Caudrey (Springer, Berlin,1980)
[11] M J Alowitz, A Ramani and H Segur, *J. Math. Phys.* **27**, 1006 (1980)
[12] J Weiss, M Tabor and G Carnevale, *J. Math. Phys.* **24**, 532 (1983)
[13] M D Kruskal, Private Communication (1982)
[14] J Weiss, *J. Math. Phys.* **24**, (1983), 1405, 532; **25**, (1984), 13,2226; **26**, (1985), 258, 2174 (1986)
[15] A C Newell, M Tabor and Y B Zeng, *Physica* **D29**, 1 (1987)
[16] F Lund and T Regge, *Phys. Rev.* **D14**, 1524 (1976)
[17] K Pohlymeyer, *Com. Math. Phys.* **46**, 207 (1977)
[18] B S Getmanov, JINR Preprint E2-11093, Dubna 1977
[19] K Porsezian, *Chaos, solitons and fractals* (Submitted)
[20] P Goldstein and E Infeld, *Phys. Lett.* **A103**, 8 (1984)
[21] K Porsezian and K Nakkeeran, *J. Mod. Opt.* **42**, 1953 (1995)
[22] K Porsezian and K Nakkeeran, *Phys. Rev. Lett.* **76**, 3955 (1996)
[23] S Liu and W Z Wang, *Phys. Rev.* **E49**, 5726 (1994)
[24] K Porsezian, *Pure. App. Opt.* **5**, L345 (1996)
[25] F Cariello and M Tabor, *Physica* **D39**, 77 (1989)
[26] M Lakshmanan and P Kaliappan, *J. Math. Phys.* **24**, 795 (1983)
[27] J D Gibbon and M Tabor, *J. Math. Phys.* **26**, 1956 (1985)
[28] R Conte, A P Fordy and A Pickering, *Physica* **D69**, 33 (1993)
[29] R Conte, *Phys. Lett.* **A140**, 383 (1989)
[30] A Pickering, *J. Math. Phys.* **37**, 1894 (1996)
[31] F Calogero and A Degasperis, *Lett. Nuovo Cimento* **22**, 139, 271 421 (1978)
[32] R Balakrishnan, *Phys. Rev.* **A32**, 1144 (1985)
[33] H H Chen and C S Liu, *Phys. Fluids* **21**, 377 (1978)
[34] S P Burtsev, V E Zakharov and A V Mikhailov, *Teor. Math. Fiz.* **70**, 323 (1987)
[35] M Daniel, K Porsezian and M Lakshmanan, *J. Math. Phys.* **35**, 6498 (1994)
[36] K Porsezian and K Nakkeeran, *Phys. Lett.* **A206**, 183 (1995)
[37] K Porsezian, M Daniel and R Bharathikannan, *Phys. Lett.* **A156**, 206 (1991)
[38] K Porsezian and M Lakshmanan, *J. Math. Phys.* **32**, 2923 (1992)
[39] K Porsezian, *Phys. Lett.* **A191**, 229 (1994)
[40] K Porsezian, *Phys. Lett.* **A221**, 163 (1996)
[41] W H Steeb and J A Louw, *Phys. Lett.* **A113**, 61 (1985)
[42] S Puri, *Phys. Lett.* **A107**, 359 (1985)
[43] N Joshi, *Phys. Lett.* **A125**, 456 (1987)
[44] L Halavaty, *J. Phys. Soc. Japan* **55**, 1405 (1995); *Phys. Lett.* **A128**, 335 (1988)
[45] W H Steeb, M Kloke, B M Spieker and W Oevel, *J. Phys.* **A16**, L447 (1983)
[46] N Nirmala, M J Vedan and B V Baby, *J. Math. Phys.* **27**, 2640, 2644 (1986)
P A Clarkson, *Proc. R. Soc. Edinb.* **A109**, 109 (1988)

[47] R Radha and M Lakshmanan, *J. Phys.* **A28**, 6977 (1995)
[48] K Porsezian and A Uthyakumar, *Phys. Lett.* **A183**, 371 (1993)
[49] W Oevel and W H Steeb, *Phys. Lett.* **A103**, 239 (1984)
[50] S Ganesan and M Lakshmanan, *J. Phys.* **A20**, L1147 (1988)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 1,
January 1997
pp. 163–188

Localized coherent structures of $(2+1)$ dimensional generalizations of soliton systems

M LAKSHMANAN and R RADHA
Centre for Nonlinear Dynamics, Department of Physics, Bharathidasan University,
Tiruchirapalli 620 024, India

Abstract. We briefly review the recent progress in obtaining $(2+1)$ dimensional integrable generalizations of soliton equations in $(1+1)$ dimensions. Then, we develop an algorithmic procedure to obtain interesting classes of solutions to these systems. In particular using a Painlevé singularity structure analysis approach, we investigate their integrability properties and obtain their appropriate Hirota bilinearized forms. We identify line solitons and from which we introduce the concept of ghost solitons, which are patently boundary effects characteristic of these $(2+1)$ dimensional integrable systems. Generalizing these solutions, we obtain exponentially localized solutions, namely the dromions which are driven by the boundaries. We also point out the interesting possibility that while the physical field itself may not be localized, either the potential or composite fields may get localized. Finally,the possibility of generating an even wider class of localized solutions is hinted by using curved solitons.

Keywords. Solitons in higher dimensions; integrability; coherent structures; Painlevé analysis; Hirota method.

PACS Nos 03.40; 11.10; 03.50

1. Introduction

The concept of soliton was born from the now famous numerical experiments of Zabusky and Kruskal [1] on the stability properties of solitary wave solutions of the Korteweg-de Vries (KdV) equation. These results were then analytically confirmed by the seminal work of Gardner, Greene, Kruskal and Miura [2], who essentially developed the method of inverse scattering transform (IST) procedure to solve the Cauchy initial value problem of the KdV equation. Since then,a very large class of $(1+1)$ dimensional nonlinear dispersive wave equations such as the sine-Gordon (sG), nonlinear Schrödinger (NLS), Heisenberg ferromagnetic spin equations and so on have been identified [3, 4] which are also solvable by the IST method exhibiting the so called soliton solutions. Furthermore,these systems are understood to be completely integrable infinite dimensional dynamical systems. These equations encompass a vast number of topics in $(1+1)$ dimensions in many areas of physics and other sciences and engineering [5]. In $(1+1)$ dimensions, the soliton systems admit solitary waves and these waves retain their shape and speed under collisions. Mathematically, the corresponding nonlinear evolution equations are linearizable and linear spectral problems can be identified with each one of them. What are the possible counterparts of such soliton systems in 2 space and 1 time, structures can be identified here, whenever appropriate boundary effects are included in the $(2+1)$ dimensional systems. The plan of the review is as follows. In § 2, we give a very brief sketch of the concept of solitons in $(1+1)$ dimensions and some of the counterparts in $(2+1)$ dimensions. In § 3, we point out how such soliton equations are associated with linear spectral problems and how their initial value problems can be solved using the inverse scattering transform method. In addition, we point out how by treating the IST problem as just a d-bar problem associated with analytic functions, the above procedure can be extended to $(2+1)$ dimensions. In addition, we list some of the important $(2+1)$ dimensional integrable nonlinear evolution equations which have been considered in the recent literature and possible algorithmic procedures to look for exponentially localized solutions. In § 4, we sketch the salient features associated with the Painlevé analysis and Hirota bilinearization procedure treating KdV as an example. Exponentially localized solutions for the simplest $(2+1)$ dimensional generalization of KdV is obtained in § 5 indicating the influence of boundary effects. Section 6 contains brief details of the results for the other $(2+1)$ dimensional evolution equations. Finally § 7 gives a summary of the results.

2. The concept of soliton and its generalization

2.1 *Solitons in* $(1+1)$ *dimensions*

For linear dispersive systems, it is well known that the associated evolution equation admits elementary wave solutions of trignometric type, satisfying appropriate dispersion relations $\omega = \omega(k)$. The Fourier transform of such elementary solutions over the k-space then gives the general solution, thereby solving the initial value problem [6].

In contrast, often the elementary wave solutions of nonlinear dispersive systems are expressible in terms of elliptic functions (example cnoidal waves) and in particular, one has amplitude-dependent dispersion relations, $\omega = \omega(k, A)$. For appropriate choice of constants, one can often obtain solitary waves as special solutions for these wave equations. However, because of the nonlinearity, Fourier transform of such elementary solutions will no longer be a solution. On the other hand,one may be able to identify a nonlinear superposition principle asympotically.

Example: Korteweg – de Vries equation

$$u_t + 6uu_x + u_{xxx} = 0. \tag{1}$$

The solitary wave solution is

$$u = u_1(x,t) = \frac{k^2}{2}\mathrm{sech}^2\left[\frac{k}{2}(x - k^2 t + \delta)\right], \tag{2}$$

where k and δ are constants. One can also have a more general solution

$$u = u_2(x,t) = 2(k_2^2 - k_1^2)\left[\frac{k_2^2\,\mathrm{cosech}^2\chi_2 + k_1^2\,\mathrm{sech}^2\chi_1}{(k_2 \coth\chi_2 - k_1 \tanh\chi_1)^2}\right], \tag{3}$$

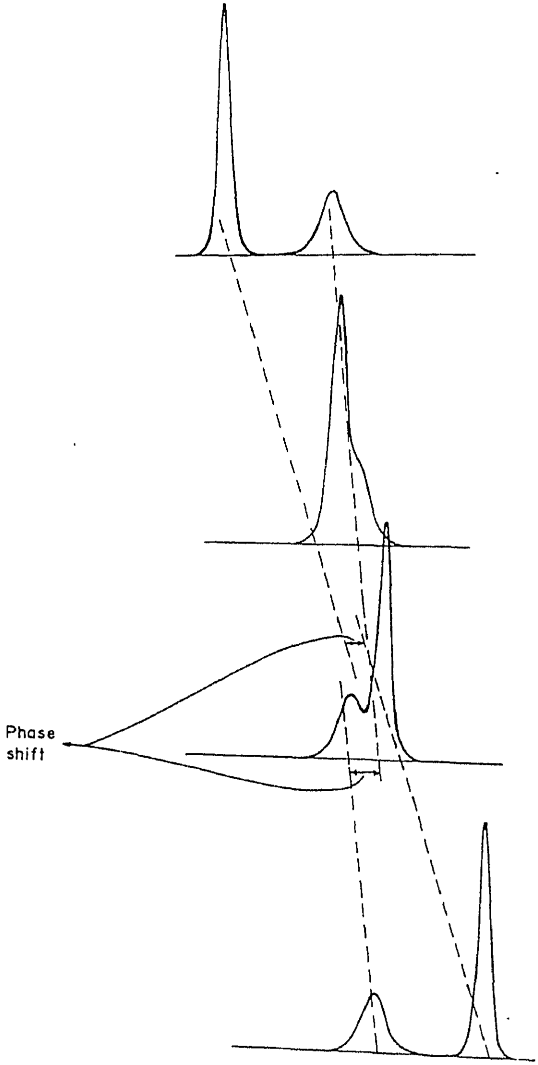


Figure 1. Two soliton interaction in the KdV equation (see (3)).

$k_i x - 4k_i^3 t + \delta_i$, $i = 1, 2$ (k_i, δ_i: constants). With a little bit of analysis, one can check that

$$u_2(x,t) \xrightarrow[t \to -\infty]{} \begin{cases} 2k_1^2 \operatorname{sech}^2(\chi_1 - \Delta) & \text{as} \quad \chi_1 \to \mathbf{O}(1), \chi_2 \to \infty \\ 2k_2^2 \operatorname{sech}^2(\chi_2 + \Delta) & \text{as} \quad \chi_2 \to \mathbf{O}(1), \chi_1 \to \infty \end{cases} \tag{4}$$

$$u_2(x,t) \xrightarrow[t \to \infty]{} \begin{cases} 2k_1^2 \operatorname{sech}^2(\chi_1 + \Delta) & \text{as} \quad \chi_1 \to \mathbf{O}(1), \chi_2 \to \infty \\ 2k_2^2 \operatorname{sech}^2(\chi_2 - \Delta) & \text{as} \quad \chi_2 \to \mathbf{O}(1), \chi_1 \to \infty \end{cases} \tag{5}$$

and so one can interpret the solution (3) as representing two solitary waves of differing amplitudes travelling in the same direction undergoing elastic collisions and coming out without change of shape or velocity except for a phase shift. The actual plot of the solution (3) for different times as given in figure 1 actually confirms this interpretation. Such solitary waves are then essentially called solitons. In fact,one can obtain explicit N-soliton solutions too as is well known [3] and also show that the solution of the Cauchy initial value problem for general initial conditions leads asymptotically to such N-soliton solutions in the background of small oscillatory decaying waves.

The above soliton picture holds good not only for the KdV equation, but for many other ubiquitous systems in $(1+1)$ dimensions; sine-Gordon, nonlinear Schrödinger, modified KdV, Heisenberg continuum spin chain equations are some of the standard examples [3].

2.2 *Excitations in* $(2+1)$ *dimensions*

A question naturally arises as to what are the basic nonlinear excitations in higher dimensions, especially in $(2+1)$ dimensions. In recent years, at least three basic excitations have been identified in $(2+1)$ dimensional nonlinear generalizations of $(1+1)$ dimensional soliton possessing systems. They are

(i) line solitons (decaying everywhere except along certain lines),
(ii) lump solitons (algebraically decaying) and
(iii) dromions (exponentially decaying).

The line solitons are nothing but straightforward generalization of the soliton solutions to $(2+1)$ dimensions. They decay exponentially everywhere in space except along certain lines, where they are bounded. For example, for the case of the so called Kadomtsev–Petviashvilli I (KPI) equation,

$$(u_t + 6uu_x + u_{xxx})_x - 3u_{yy} = 0, \tag{6}$$

which is obviously a generalization of KdV, one has the line soliton solution (figure 2)

$$u(x,y,t) = \frac{k^2}{2}\,\mathrm{sech}^2[\tfrac{1}{2}(kx + 2k^3t + k^2y + \delta)]. \tag{7}$$

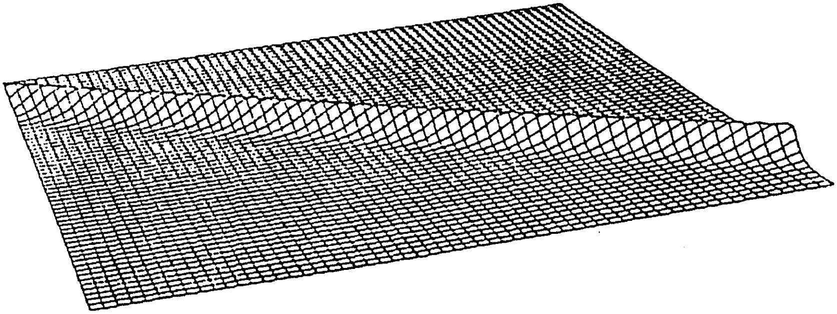


Figure 2. Line soliton solution of the KPI equation (see (7)).

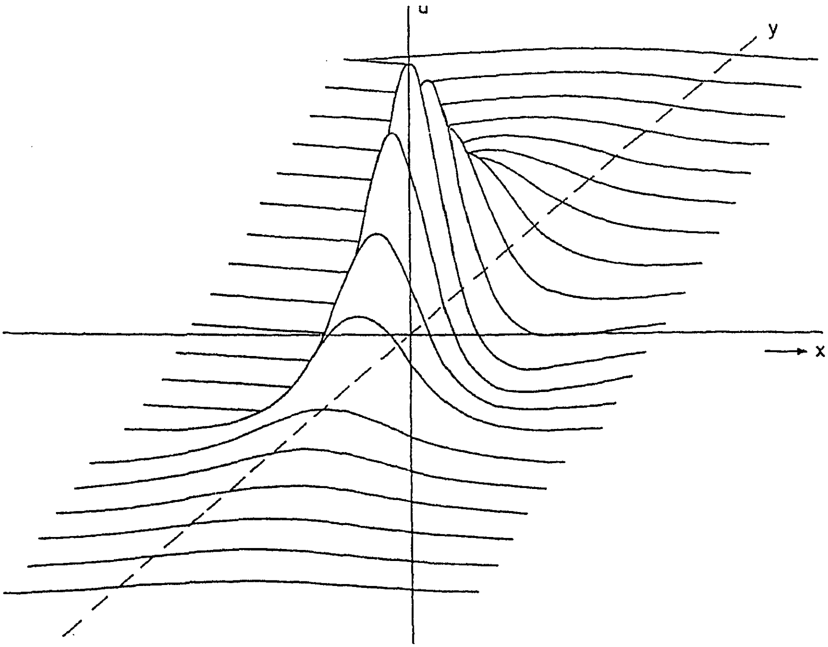


Figure 3. Lump soliton solution of the KPI equation (see (8)).

Equation (6) has also a lump soliton of the form

$$
\begin{aligned}
u &= \frac{4[-(x' + k_R y')^2 + k_I^2 y'^2 - 3/k_I^2]}{[(x' + k_R y')^2 + k_I^2 y'^2 - 3/k_I^2]^2}, \\
x' &= x + (k_R^2 + k_I^2)t, \\
y' &= \frac{1}{\sqrt{3}} y - 2k_R t,
\end{aligned} \tag{8}
$$

which decays algebraically and is illustrated in figure 3. Finally, considering the Davey – Stewartson equation I,

$$
\begin{aligned}
& iq_t + q_{\xi\xi} + q_{\eta\eta} + (\nu_1 + \nu_2)q = 0, \\
& \nu_{1\xi} = -2\epsilon |q|_\eta^2, \\
& \nu_{2\eta} = -2\epsilon |q|_\xi^2, \quad \epsilon = \pm 1,
\end{aligned} \tag{9}
$$

it admits exponentially localized solutions [7, 8], namely dromions which are driven along specified tracks determined by the boundaries (dromos in Greek) of the form

$$
q(\xi, \eta, t) = \frac{\rho \exp\{p_{1R}\hat{\xi} + s_{1R}\hat{\eta} + i[p_{1I}\hat{\xi} + s_{1I}\hat{\eta} + \{|p_1|^2 + |s_1|^2\}t]\}}{1 + j\exp(2p_{1R}\hat{\xi}) + k\exp(2s_{1R}\hat{\eta}) + l\exp(2(p_{1R}\hat{\xi} + s_{1R}\hat{\eta}))}, \tag{10a}
$$

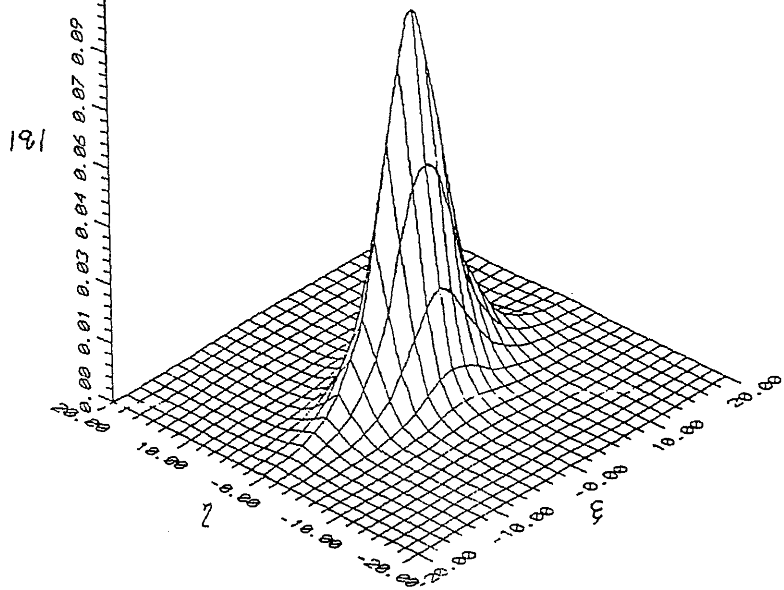


Figure 4. (1,1) dromion solution of the DSI equation (see (10)).

where

$$\hat{\xi} = \xi - 2p_{1I}t, \quad \hat{\eta} = \eta - 2s_{1I}t \tag{10b}$$

and p, s, j, k, l are constants. The form of (10a) is explicitly shown in figure 4.

In the remaining sections, we wish to explain how such exponentially localized dromion possessing equations can be generated and how such solutions may be realized. To start with, we demonstrate this for the KdV type equations.

3. Linearization and IST-(2 + 1) dimensional KdV

It is well known [3] that the $(1+1)$ dimensional soliton systems are linearizable in the sense that given the nonlinear evolution equation

$$u_t + K(u) = 0, \tag{11}$$

where K is some nonlinear functional, then the compatibility of the two linear systems

$$L_1\psi = 0, \quad L_2\psi = 0, \tag{12}$$

where L_1 and L_2 are linear differential operators, gives rise to the Lax equation

$$[L_1, L_2] = 0, \tag{13}$$

which is equivalent to the evolution equation (11). Example: KdV:

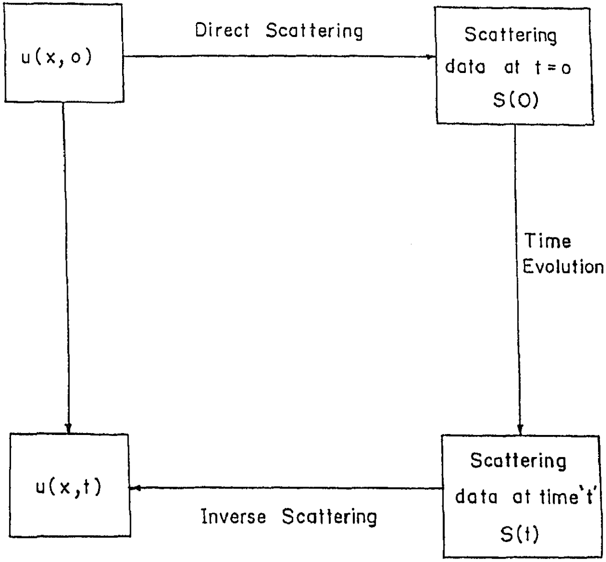


Figure 5. The schematic representation of the IST method.

$$L_1 = -\frac{\partial^2}{\partial x^2} + (u(x,t) - \lambda), \tag{14a}$$

(Schrödinger operator)

$$L_2 = \frac{\partial}{\partial t} + 4\frac{\partial^3}{\partial x^3} + 6u\frac{\partial}{\partial x} + 3u_x. \tag{14b}$$

Then given the initial data $u(x,0)$ with $u \underset{|x|\to\infty}{\longrightarrow} 0$, (i) analyzing the scattering problem for $L_1\psi = 0$, the scattering data $S(0)$ at time $t = 0$ can be obtained. (ii) Making use of $L_2\psi = 0$, one can obtain the time evolution of $S(t)$ almost trivially. (iii) Carrying out an inverse scattering transform procedure for the scattering data $S(t)$ and by solving the associated Gelfand–Levitan–Marchenko equation, the 'potential' $u(x,t)$ can be uniquely determined and so the initial value problem can be solved.

The above three steps are schematically illustrated in figure 5. Now the crucial point is that an extension of the above procedure to $(2+1)$ dimensions can also be carried out provided we interpret the IST procedure as (i) a d-bar problem in complex analysis for the direct scattering and (ii) the inverse scattering part as equivalent to obtaining the analytic function (given the d-bar data) using the generalized Cauchy integral formula [9–11].

Thus the extension of the IST analysis crucially depends upon the linearization of $(2+1)$ dimensional equations and identifying appropriate Lax pairs in order to carry out the d-bar analysis. One way of doing this is to generalize the Lax pairs L_1 and L_2 as given in (12) by introducing operators and functions dependent on a second spatial variable y in a consistent way such that the Lax equation (13) gives the appropriate nonlinear evolution equation. We demonstrate this for the case of KdV equation in the following subsection. Similar analysis can be carried out for other ubiquitous soliton equations [12].

3.1 *Generalization of KdV to* $(2+1)$ *dimensions*

There are several ways to generalize the Lax operators L_1 and L_2 but retaining the form of the Lax equation (13). We may identify at least three interesting possibilities in the case of the KdV equation. They are as follows.

Method 1 [13]. One essentially includes first derivative operators of y in L_1 while making compensatory changes in L_2 in order to obtain the evolution equation from Lax condition. For example for the KdV Lax pair (14), one can make the modifications

$$L_1 = \sigma\partial_y + \partial_x^2 + u(x,y,t) \tag{15a}$$

and

$$L_2 = \partial_t + 4\partial_x^3 + 6u\partial_x + 3u_x + 3\sigma\partial_x^{-1}u_y \tag{15b}$$

so as to get the KPI and KPII equations (for $\sigma = i$ and 1 respectively).

Method 2 [14]. One keeps the operator L_1 unchanged while introducing first derivatives of y in L_2:

$$L_1 = L(\partial_x) - \lambda, \tag{16a}$$

$$L_2 = \partial_t + f(\lambda)\partial_y + \sum_{m=0}^{\mu} \nu_m(x,y,t)\partial_x^m. \tag{16b}$$

For KdV, this reads

$$\begin{aligned} L_1 &= -\partial_x^2 + u(x,y,t) - \lambda, \\ L_2 &= \partial_t - 4\beta\partial_x^2\partial_y + 4\beta u_x\partial_y + 3\beta u_{xy} + 2\beta u\partial_x. \end{aligned} \tag{17}$$

Then the Lax equation $[L_1, L_2] = 0$ leads to the so called breaking soliton equation:

$$u_t = 4\beta uu_y + 2\beta u_x\partial_x^{-1}u_y - \beta u_{xxy} = 0, \tag{18}$$

provided the eigenvalue λ in (17) evolves as

$$\lambda_t = \lambda\lambda_y \tag{19}$$

leading to the breaking wave nature in the y direction.

Method 3 [15]. One can also look at the possibility of introducing derivatives in the second (y) variable with order greater than one. However such a possibility in general leads to constant coefficients of L_1 or L_2 and so to trivial evolution equations. To circumvent this difficulty in the case of $(2+1)$ dimensional systems, one realizes that it is enough to have only a sufficiently broad class of solutions of the linear equations $L_1\psi = 0$ and $L_2\psi = 0$ and that the spectrum need not be common. As a result, one can impose the weaker Lax condition

$$[L_1, L_2]\psi = 0 \tag{20}$$

in the subspace of the eigenfunction ψ or equivalently

$$[L_1, L_2] = \gamma L_1, \tag{21}$$

where γ is some operator.

Example: KdV generalization

$$L_1 = (\partial_x^2 - \sigma^2 \partial_y^2) + u(x,y,t) = \partial_\xi \partial_\eta + u(\xi,\eta,t), \tag{22}$$

$$L_2 = \partial_t + k_1 \partial_\xi^3 + k_1 \partial_\eta^3 + 3k_1 \partial_\xi^{-1} u_\eta \partial_\eta + 3k_1 \partial_\eta^{-1} u_\xi \partial_\xi. \tag{23}$$

From the weak Lax equation, one can then obtain the Boiti *et al* generalized KdV equation [15]

$$u_t + k_1 u_{\xi\xi\xi} = 3k_1 (u \partial_\eta^{-1} u_\xi)_\xi \tag{24}$$

when $k_2 = 0$ in (23). When $k_2 \neq 0$, the Nizhnik–Novikov–Veselov (NNV) equation is obtained as

$$u_t + k_1 u_{\xi\xi\xi} + k_2 u_{\eta\eta\eta} = 3k_2 (u \partial_\xi^{-1} u_\eta)_\eta + 3k_1 (u \partial_\eta^{-1} u_\xi)_\xi. \tag{25}$$

For a more general class of equations, see for examaple Ablowitz and Clarkson [3]. In this way,the other soliton equations in $(1+1)$ dimensions can also be generalized to $(2+1)$ dimensions though the exact procedure may vary from system to system [3]. Some of the higher dimensional systems which have been studied extensively in recent times include the following evolution equations apart from the ubiquitous K–P and Davey–Stewartson equations already mentioned:

(1) Generalized KdV [15]

$$u_t + u_{\xi\xi\xi} = 3(u \partial_\eta^{-1} u_\xi)_\xi. \tag{26}$$

(2) Generalized Nizhnik–Novikov–Veselov (NNV) equation[16]

$$u_t + u_{\xi\xi\xi} + u_{\eta\eta\eta} + a u_\xi + b u_\eta = 3(u \partial_\eta^{-1} u_\xi)_\xi + 3(u \partial_\xi^{-1} u_\eta)_\eta. \tag{27}$$

(3) Breaking soliton equation [12]

$$u_t + a u_{xxx} + 6 a u u_x + b u_{xxy} + 4 b u u_y + 4 b u_x \partial_x^{-1} u_y = 0. \tag{28}$$

(4) $(2+1)$ dimensional generalized NLS equation [17]

$$iq_t - (\alpha - \beta) q_{\xi\xi} + (\alpha + \beta) q_{\eta\eta} - 2\lambda q \left[(\alpha + \beta) \left\{ \int_{-\infty}^{\xi} |q|_\eta^2 \mathrm{d}\xi' + u_1(\eta, t) \right\} \right.$$
$$\left. - (\alpha - \beta) \left\{ \int_{-\infty}^{\eta} |q|_\xi^2 \mathrm{d}\eta' + u_2(\xi, t) \right\} \right] = 0. \tag{29}$$

(5) $(2+1)$ dimensional generalized NLS equation [18]

$$iq_t = q_{xy} + Vq, V_x = 2\partial_y |q|^2. \tag{30}$$

(6) (2 + 1) dimensional generalized sine-Gordon equation [19, 20]

$$\theta_{\xi\eta t} + \tfrac{1}{2}\theta_\eta \rho_\xi + \tfrac{1}{2}\theta_\xi \rho_\eta = 0, \tag{31a}$$

$$\rho_{\xi\eta} = \tfrac{1}{2}(\theta_\xi \theta_\eta)_t. \tag{31b}$$

(7) (2 + 1) dimensional long dispersive wave (2LDW) equation [21]

$$\lambda q_t + q_{xx} - 2qv = 0, \tag{32a}$$

$$\lambda r_t - r_{xx} + 2rv = 0, \tag{32b}$$

$$(qr)_x = v_\eta, \partial_\eta = \partial_x - \lambda \partial_y. \tag{32c}$$

We will consider the existence of exponentially localized solutions in all these systems in the next sections.

4. Painlevé analysis and Hirota bilinearization

4.1 *Painlevé analysis*

In order to investigate the integrability aspects of the above type of nonlinear evolution equations, one can proceed to analyze the singularity structure of the solutions in the complex space of the independent variables, namely the so called Painlevé property. Such a procedure has been found to give considerable wealth of informations for finite dimensional nonlinear systems [22–24]. Application to pdes has also been well developed and the procedure of Weiss, Tabor and Carnevale [25] is algorithmic and identifies the nature of the singularities in the local neighbourhood of a noncharacteristic singular manifold.

Let us consider a NLEE of the form

$$u_t + K(u) = 0, \tag{33}$$

where $K(u)$ is a nonlinear functional of $u(x_1, x_2, \ldots, x_m, t) = u(X, t)$ and its derivatives upto order N so that (33) is an Nth order nonlinear pde. Then one can say that (33) possessess the P-property when the following conditions are satisfied.

The solution of (33) must be single valued about the noncharacteristic movable singular manifold. More precisely, if the singular manifold is determined by

$$\phi(x,t) = 0, \quad \phi_{x_i}(x,t) \neq 0, \quad \phi_t(x,t) \neq 0, \quad i = 1, \ldots, M \tag{34}$$

and $u(x,t)$ is a solution of the pde (33), then we have the Laurent expansion

$$u(x,t) = [\phi(x,t)]^{-m} \sum_{j=0}^{\infty} u_j(x,t) \phi^j(x,t) \tag{35}$$

in a deleted neighbourhood of the singular manifold (34) and m is an integer. By the Cauchy–Kovalevskaya theorem, the solution (35) should contain N arbitrary functions, one of them being the singular manifold ϕ itself and the others coming from the u_j's.

Then the WTC procedure to test the given pde for its P-property essentially consists of the following three steps [25].

(i) Determination of leading order behaviours.

(ii) Identification of powers j (resonance) at which the arbitrary functions can enter into the Laurent series expansion (35) and

(iii) Verifying that at the resonance values j, a sufficient number of arbitrary functions exist without the introduction of a movable critical singular manifold. An important feature of the WTC formalism is that the generalized Laurent series expansion can not only reveal the singularity structure aspects of the solution and integrability nature of a given pde, but can also provide an effective algorithm which in most cases successfully captures all its properties, namely the linearization (Lax pair), the Bäcklund transformation (BT), Hirota bilinearization, symmetries and so on [3]. As a simple application, we illustrate the above aspects with KdV as an example. Any other soliton system can also be likewise analysed [3, 26].

For the KdV equation (1), we substitute the formal Laurent expansion (35) around the singularity manifold $\phi(x,t)=0$ into (1) and equate the powers of ϕ to zero. One finds that the exponent $m=+2$ and that at $j=-1,4,6$ arbitrary functions can enter the Laurent series (35). Identifying the arbitrariness of ϕ with $j=-1$ recursively, one finds

$$j=0: u_0=-2\phi_x^2, \tag{36a}$$

$$j=1: u_1=2\phi_{xx}, \tag{36b}$$

$$j=2: \phi_x\phi_t+6u_2\phi_{xx}-2u_3\phi_x^2+\phi_{xxxx}=0, \tag{36c}$$

$$j=3: \phi_{xt}+6u_2\phi xx-2u_3\phi_x^2+\phi_{xxxx}=0, \tag{36d}$$

$$j=4: \partial_x(\phi_{xt}+6u_2\phi_{xx}-2u_3\phi_x^2+\phi_{xxxx})=0. \tag{36e}$$

Now, it is clear that by the condition (36d), (36e) is always satisfied so that $u_4(x,t)$ is arbitrary. Similarly, one can derive the condition at $j=5$ and prove that at $j=6$, $u_6(x,t)$ is arbitrary. As the KdV equation is of third order,the Laurent series admits three arbitrary functions and so the Painlevé property is satisfied.

Now, if the arbitrary functions u_4 and u_6 are chosen to be identically zero and if we require $u_3=0$, then $u_j=0, j>3$ provided u_2 satisfies the KdV equation. Thus, we obtain the Bäcklund transformation for the KdV in the form

$$u=2(\log\phi)_{xx}+u_2, \tag{37}$$

where u_1 and u_2 solve the KdV and ϕ satisfies (36a–e) with $u_3=0$. By a set of transformations, it is possible to show that the defining equations for ϕ are equivalent to linearizing equations (12) and (14). One can apply a similar procedure to any other NLEE in $(1+1)$ or $(2+1)$ dimensions and obtain its integrability property [24]. For recent applications in $(2+1)$ dimensions, see for example Radha and Lakshmanan [27–33].

4.2 *Hirota bilinearization*

We have noted above that the Bäcklund transformation of the form (37) can be used to bilinearize the NLEE. Indeed considering the vaccum solution $u_2=0$ in (37) one obtains the transformation

$$u = 2(\log\phi)_{xx} \tag{38}$$

so that the KdV equation can be written in the bilinear form

$$\phi_{xt}\phi - \phi_x\phi_t + \phi_{xxxx}\phi - 4\phi_{xxx}\phi_x + 3\phi_{xx}^2 = 0. \tag{39}$$

Using the so-called Hirota's bilinear operator

$$D_x^m D_t^n(a\cdot b) = (\partial_x - \partial_{x'})^m(\partial_t - \partial_{t'})^n a(x,t)b(x',t')|_{x'=x y'=y}, \tag{40}$$

equation (39) can be rewritten in a compact form as

$$(D_xD_t + D_x^4)\phi\cdot\phi = 0. \tag{41}$$

The properties of the bilinear operator can be easily worked out. For example, one has the relations $D_x^m a\cdot 1 = \partial_x^m a$, $D_x^m(a\cdot b) = (-1)^m D_x^m(b\cdot a)$, $D_x^m(a\cdot a) = 0$, m odd, and so on. Using such properties, the calculations can be simplified considerably.

Now, expanding ϕ in a formal power series in ϵ,

$$\phi = \sum_{n=0}^{\infty}\epsilon^n\phi^{(n)}, \quad \phi^{(0)} = 1, \tag{42a}$$

where

$$\phi^{(1)} = \sum_{i=1}^{N}\exp(\eta_i), \eta_i = k_ix + \omega_it + \eta_i^{(0)} \tag{42b}$$

and k_i, ω_i, $\eta_i^{(0)}$ are constants, the N soliton solution of KdV can be obtained. To see this, one substitutes (42) in (41) and equates each power of ϵ separately to zero to obtain the following set of equations:

$$O(1) : 0 = 0, \tag{43a}$$

$$O(\epsilon) : \phi_{xt}^{(1)} + \phi_{xxxx}^{(1)} = 0, \tag{43b}$$

$$O(\epsilon^2) : \phi_{xt}^{(2)} + \phi_{xxxx}^{(2)} = -\tfrac{1}{2}(D_xD_t + D_x^4)\phi^{(1)}\cdot\phi^{(1)}, \tag{43c}$$

and so on. The procedure is then to use (43b) in (43c) and succesively solve the remaining equations. In practice, one finds the solution for $N = 1,2,3$ and then hypothesize it for arbitrary N which is to be proved by induction.

For example, for $N = 1$, $\phi^{(1)} = \exp(\eta_1)$ and from (43b), $\omega_1 = -k_1^3$ and $(\partial_x\partial_t + \partial_x^4)$ $\phi^{(2)} = 0$ so that $\phi^{(2)} = 0$ and $\phi^{(i)} = 0, i \geq 2$. Thus, the solution of (43b) can be written as

$$\phi^{(1)} = \exp(\eta_1), \quad \eta_1 = k_1x - k_1^3t + \eta_1^{(0)}. \tag{44}$$

Making use of (38) and (42a), it is straightforward to obtain the one soliton solutions of KdV given by (2).

Similarly for the case $N = 2$,

$$\phi^{(1)} = \exp(\eta_1 + \eta_2), \quad \eta_i = k_ix - k_i^3t + \eta_i^{(0)}, \quad i = 1,2 \tag{45a}$$

so that the solution (42a) becomes

$$\phi^{(2)} = 1 + \exp(\eta_1) + \exp(\eta_2) + \exp(\eta_1 + \eta_2 + A_{12}), \tag{45b}$$

where A_{12} is a constant. Again this leads to the 2-soliton solution (3) of the KdV equation. In an analogous fashion, one can proceed to find the N-soliton solution also. It is also useful to note that with the solution of the Hirota equation in a form such as (45b), it is quite easy to understand the elastic nature of the soliton interaction as discussed in § 2.

As noted above, all the other known soliton equations in $(1+1)$ dimensions can also be bilinearized and the soliton solution obtained through the Hirota method. In the next sections, we will apply this method to generate interesting coherent structures in $(2+1)$ dimensions.

5. Localized coherent structures in the $(2+1)$ dimensional KdV equation

As an example,we consider the generalized KdV equation (26) of Boiti *et al* and obtain its localized solutions [28]. We rewrite it as

$$u_t + u_{\xi\xi\xi} = 3(uv)_\xi, \tag{46a}$$

$$u_\xi = v_\eta. \tag{46b}$$

Note that the potential v itself satisfies the equation

$$v_t + v_{\xi\xi\xi} = \tfrac{3}{2}(v^2)_\xi + 3\partial_\eta^{-1}[v_{\xi\xi}\partial_\xi^{-1}v_\eta + v_\xi v_\eta] \tag{47}$$

5.1 *P-property*

Considering the Laurent expansion

$$u = \sum_{j=0}^{\infty} u_j(x,y,t)\phi^{j+\alpha}, \tag{48a}$$

$$v = \sum_{j=0}^{\infty} v_j(x,y,t)\phi^{j+\beta}, \tag{48b}$$

around the noncharacteristic singular manifold $\phi(x,y,t) = 0$, ϕ_x, ϕ_y, , $\phi_t \neq 0$, we can algorithmically check that

$$\alpha = \beta = -2, \quad u_0 = 2\phi_\xi\phi_\eta, \quad v_0 = 2\phi_\xi^2 \tag{49}$$

and that one member each of (u_2, v_2), (u_4, v_4) and (u_6, v_6) are (the coefficient functions) arbitrary in addition to the manifold ϕ itself without the introduction of a critical singular manifold. Thus, the P-property is satisfied here.

5.2 *Hirota bilinear form*

One can easily check that the Laurent series (48) can be cut off at constant level terms to consider special solutions of the form

$$u = \frac{u_0}{\phi^2} + \frac{u_1}{\phi} + u_2, \tag{50a}$$

$$v=\frac{v_0}{\phi^2}+\frac{v_1}{\phi}+v_2, \tag{50b}$$

so that if (u_2,v_2) solve (46), so also (u,v). Thus, (50) can be effectively considered as an auto-Bäcklund transformation and its further properties can be worked out. Particularly useful to us is the vacuum solution

$$u_2=v_2=0. \tag{51}$$

Then with the expressions (49) for (u_0,v_0) and (u_1,v_1) (from a knowledge of the P-analysis), the transformations (50) can be simplified to

$$u=-2\partial_{\xi\eta}\log\phi, \tag{52a}$$

$$v=-2\partial_{\xi\xi}\log\phi. \tag{52b}$$

On substituting the above into the original evolution equation (46), one obtains the bilinear version

$$2\phi\phi_{\eta t}-2\phi_\eta\phi_t+2\phi\phi_{\xi\xi\xi\eta}-2\phi_{\xi\xi\xi}\phi_\eta+6\phi_{\xi\xi}\phi_{\xi\eta}-6\phi_\eta\phi_{\xi\xi\eta}=0. \tag{53}$$

In the Hirota operator notation, (53) can be rewritten as

$$(D_\eta D_t+D_\xi^3D_\eta)\phi\cdot\phi=0. \tag{54}$$

Expanding ϕ as a power series,

$$\phi=1+\varepsilon\phi^{(1)}+\varepsilon^2\phi^{(2)}+\cdots, \tag{55}$$

equation (54) can be written as a set of coupled linear pdes:

$$\mathrm{O}(\varepsilon):\quad \phi^{(1)}_{\eta t}+\phi^{(1)}_{\xi\xi\xi\eta}=0, \tag{56a}$$

$$\mathrm{O}(\varepsilon^2):\quad \phi^{(2)}_{\eta t}+\phi^{(2)}_{\xi\xi\xi\eta}=-\tfrac{1}{2}(D_\eta D_t+D_\xi^3D_\eta)\phi^{(1)}\cdot\phi^{(1)}, \tag{56b}$$

etc.

5.3 *Line solitons and ghost solitons*

Equation (56a) obviously admits special classes of solutions of the form

$$\phi^{(1)})=\sum_{j=1}^{N}\exp(\chi_j),\quad \chi_j=k_j\xi+l_j\eta-k_j^3t+\chi_j^{(0)}, \tag{57}$$

where k_j, l_j and $\chi_j^{(0)}$'s are constants. One can use these solutions for various choices of N to obtain the so called line solitons as in the case of KdV equation as discussed in § 4.2. For example, with the choice $N=1$, we have

$$\phi^{(1)}=\exp(\chi_1),\quad \chi_1=k_1\xi+l_1\eta-k_1^3t+\chi_j^{(0)} \tag{58}$$

and the corresponding solution to the evolution equation becomes

$$u=-2\partial_{\xi\eta}\log\phi=-2\partial_{\xi\eta}\log(1+\phi^{(1)})=-\frac{k_1l_1}{2}\operatorname{sech}^2\left(\frac{\chi_1}{2}\right), \tag{59a}$$

$$v=-2\partial_{\xi\xi}\log\phi=-2\partial_{\xi\xi}\log(1+\phi^{(1)})=-\frac{k_1^2}{2}\operatorname{sech}^2\left(\frac{\chi_1}{2}\right), \tag{59b}$$

where ε has been scaled out in the final form of the solution. These structures in $(2+1)$ dimensions are line solitons, because eventhough they have soliton-like properties, they are localized in the two dimensional space except along the lines $\chi_1 = 0$, where the solution is though bounded does not go to zero as $x, y \to \infty$. N-line solitons can also be constructed following the procedure suggested in § 4 and their interaction properties can be studied in the usual way.

But more interestingly, we observe the fact that in (59), k_1 and l_1 are arbitrary parameters. One finds that

$$\text{as} \quad k_1 \to 0, \quad \text{both} \quad u, v \to 0, \tag{60a}$$

whereas

$$\text{as} \quad l_1 \to 0,\, u \to 0,\, v \to \frac{-k_1^2}{2} \operatorname{sech}^2(k_1\xi - k_1^3 t + \chi_1^{(0)}). \tag{60b}$$

Thus eventhough the physical field vanishes, the potential

$$v = \int_{-\infty}^{\eta} u_\xi d\eta' + v_1(\xi, t) \tag{61}$$

survives. Comparing (60b) and (61), one concludes that

$$\lim_{\eta \to -\infty} v(\xi, \eta, t) = v_1(\xi, t) = -\frac{k_1^2}{2} \operatorname{sech}(k_1\xi - k_1^3 t + \chi_1^{(0)}) \tag{62}$$

is the ghost or background soliton at the boundary $\eta = -\infty$.

5.4 *Localized solution: Dromions*

The existence of ghost solitons at the boundary $\eta = -\infty$ prompts one to look for a more general ansatz than (58) so that we may choose

$$\phi = 1 + e^{\chi_1} + e^{\chi_2} + K e^{\chi_1 + \chi_2}, \quad K > 0, \tag{63a}$$

where

$$\chi_1 = k_1\xi - k_1^3 t + c_1, \quad \chi_2 = l_1\eta + c_2 \tag{63b}$$

so that

$$u = -2\partial_{\xi\eta} \log \phi = \frac{2k_1 l_1 (1-K)\exp(\chi_1 + \chi_2)}{[1 + \exp(\chi_1) + \exp(\chi_2) + K\exp(\chi_1 + \chi_2)]^2}, \tag{64}$$

while

$$v = -2\partial_{\xi\xi} \log \phi = -\frac{2k_1^2[\exp(\chi_1) + (1-K)\exp(\chi_1 + \chi_2) + \exp(\chi_1 + 2\chi_2)]}{[1 + \exp(\chi_1) + \exp(\chi_2) + K\exp(\chi_1 + \chi_2)]^2}. \tag{65}$$

One observes that while the physical field is now exponentially localized in the entire

(ξ,η) plane, the potential v is not. In this case, when

$$u \to 0 \quad \text{as} \quad \eta \to -\infty, \tag{66a}$$

$$v \to -\frac{k_1}{2}\,\mathrm{sech}^2(k_1\xi - k_1^3 t + c_1) \tag{66b}$$

so that we can interpret that the energy is continuously injected at the boundary $\eta \to -\infty$ in the form of a soliton (with $k \neq 0$) so that the physical field can overcome the tendency to disperse so as to get a fully localized structure. Solution (64) is then called the (1,1) dromion solution to the NLEE (46).

Multidromion solutions can also be obtained. For example a (2, 1) dromion solution is obtained as

$$u = \frac{2(1-K)[k_1 l_1 \exp(\chi_1+\chi_2) + k_2 l_1 \exp(\chi_2+\chi_3)]}{[1+\exp(\chi_1)+\exp(\chi_2)+K(\exp(\chi_1+\chi_2)+\exp(\chi_2+\chi_3))]^2}, \tag{67a}$$

where

$$\chi_1 = k_1\xi - k_1^3 t + \chi_1^{(0)}, \tag{67b}$$

$$\chi_2 = l_1\eta + \chi_2^{(0)}, \tag{67c}$$

$$\chi_3 = k_2\xi - k_2^3 t + \chi_1^{(0)}. \tag{67d}$$

One may obtain (N,M) dromions also proceeding in this way.

5.5 *Further general localized structures*

Unlike the case of $(1+1)$ dimensional soliton equations, the linear set of Hirota equations (56) can also have more general solutions than the one given by (57). For example, (56a) admits general solutions of the form

$$\phi = \sum_{j=1}^{N} \exp[k_j\xi - k_j^3 t + m_j(\eta)], \tag{68}$$

where $m_j(\eta)$ is an arbitrary function of η. For example with the choice $N = 1$ in (68), we can obtain

$$u(\xi,\eta,t) = -\tfrac{1}{2}k_1 m_1'(\eta)\,\mathrm{sech}^2\tfrac{1}{2}[k_1\xi - k_1^3 t + m_1(\eta)] \tag{69}$$

where $m_1'(\eta) = (\mathrm{d}m_1/\mathrm{d}\eta)$ is arbitrary (as $m_1(\eta)$ is arbitrary). Choosing $m_1'(\eta)$ suitably, one can obtain various types of localized solutions. For example, with the choice,

$$m_1'(\eta) = \mathrm{sech}^2(\alpha\eta+\beta), \quad \alpha,\beta : \text{constants} \tag{70a}$$

we have the exponentially localized solution

$$u = \gamma\,\mathrm{sech}^2(\alpha\eta+\beta)\,\mathrm{sech}^2\tfrac{1}{2}[k_1\xi - k_1^3 t + m_1(\eta)] \tag{70b}$$

Many other interesting classes of exponentially localized solutions including oscillatory and rational functions can be similarly constructed for suitable choices of $m_j(\eta)$ [35].

6. Exponentially localized structures in other (2+1) dimensional NLEEs [34]

The procedure elucidated in § 5 for the $(2+1)$ dimensional generalized KdV equation can be extended to other $(2+1)$ dimensional systems discussed in § 3 as well. In the following, we summarize the main results.

6.1 *NNV equation*

Proceeding as in the case of the generalized KdV equation (46), we rewrite (27) as

$$u_t + u_{\xi\xi\xi} + u_{\eta\eta\eta} + au_\xi + bu_\eta = 3(uv)_\xi + 3(uq)_\eta, \tag{71a}$$

$$u_\xi = v_\eta, \quad u_\eta = q_\xi, \tag{71b}$$

the Painlevé property can be established [28]. Working out the auto-Bäcklund transformation, the bilinearizing transformation becomes

$$u = -2\partial_{\xi\eta} \log \phi, \tag{72a}$$

$$v = -2\partial_{\xi\xi} \log \phi, \tag{72b}$$

$$q = -2\partial_{\eta\eta} \log \phi. \tag{72c}$$

The line solitons work out to be

$$u = -\frac{k_1 l_1}{2} \operatorname{sech}^2 \frac{\chi_1}{2}, \tag{73a}$$

$$v = -\frac{k_1^2}{2} \operatorname{sech}^2 \frac{\chi_1}{2}, \tag{73b}$$

$$q = -\frac{l_1^2}{2} \operatorname{sech}^2 \frac{\chi_1}{2}, \tag{73c}$$

where

$$\chi_1 = k_1\xi + l_1\eta - (ck_1 + k_1^3)t - (\mathrm{d}l_1 + l_1^3)t + \chi_1^{(0)} \tag{73d}$$

so that in the limit $k_1 \to 0$, u, $v \to 0$, while q survives, whereas when $l_1 \to 0$, u, $q \to 0$, but v survives. Thus we have two nonparallel ghost solitons in the present case. Following the procedure discussed in the previous section, we take ϕ to be of the same form as (63a) so that the (1,1) dromion in the present case becomes

$$u(\xi, \eta, t) = \frac{2k_1 l_1 (1-K) \exp(\chi_1 + \chi_2)}{[1 + \exp(\chi_1) + \exp(\chi_2) + K \exp(\chi_1 + \chi_2)]^2}, \quad K > 0 \tag{74}$$

where now

$$\chi_1 = k_1\xi - (ck_1 + k_1^3)t + \chi_1^{(0)}, \quad \chi_2 = l_1\xi - (\mathrm{d}l_1 + l_1^3)t + \chi_2^{(0)}. \tag{75}$$

Multidromions solutions can be similarly worked out.

6.2 *Breaking soliton equation*

Rewriting (28) as

$$u_t + au_{xxx} + bu_{xxy} + 6auu_x + 4buu_y + 4bu_x v = 0, \tag{76a}$$

$$u_y = v_x, \tag{76b}$$

one can check algorithmically that the Painlevé property is satisfied in the nontrivial case only for the choice $a = 0$, $b \neq 0$ ($b = 0$, $a \neq 0$ is the trivial KdV case) so that we have the equation of motion

$$u_t + bu_{xxy} + 4buu_y + 4bu_x v = 0, \tag{77a}$$

$$u_y = v_x. \tag{77b}$$

Note that v itself satisfies an equation of the form

$$v_t + bv_{xxy} + 4bvv_x + 4bv_y\partial_y^{-1}v_x = 0. \tag{78}$$

One verifies that bilinearization is possible only for the special case (77) and the required transformation works out to be

$$u = \tfrac{3}{2}\partial_{xx}\log\phi, \tag{79a}$$

$$v = \tfrac{3}{2}\partial_{xy}\log\phi. \tag{79b}$$

The bilinearized equation becomes

$$(D_x D_t + bD_x^3 D_y)\phi\cdot\phi = 0. \tag{80}$$

Expanding ϕ in a formal power series as before and truncating, one obtains the line soliton as

$$u = \tfrac{3}{8}k_1^2\operatorname{sech}^2\left(\frac{\chi_1}{2}\right), \tag{81a}$$

$$v = \tfrac{3}{8}k_1 l_1\operatorname{sech}^2\left(\frac{\chi_1}{2}\right), \tag{81b}$$

where

$$\chi_1 = k_1 x + l_1 y - bk_1^2 l_1 t + \chi_1^{(0)}. \tag{81c}$$

As $k \to 0$, both u and $v \to 0$, while when $l_1 \to 0$, the potential $v \to 0$, but the field u survives as a static ghost soliton of the form $u \to \frac{3}{8}\, k_1^2 \operatorname{sech}^2(k_1 x + \chi_1^{(0)})$ at the boundary $x = -\infty$.

The dromion solutions can be obtained in the usual way. For generating the (1,1) dromion, we take

$$\phi = 1 + e^{\chi_1} + e^{\chi_2} + Ke^{(\chi_1+\chi_2)}, \quad K > 0 \tag{82a}$$

$$\chi_1 = k_1 x + \chi_1^{(0)}, \quad \chi_2 = l_1 y - bk_1^2 l_1 t + \chi_2^{(0)} \tag{82b}$$

so that the potential v becomes fully localized exponentially,

$$v(x,y,t) = \frac{3}{2}\frac{k_1 l_1 (K-1)\exp(\chi_1+\chi_2)}{[1+\exp(\chi_1)+\exp(\chi_2)+K\exp(\chi_1+\chi_2)]^2}, \tag{83}$$

whereas the physical field

$$u(x,y,t) = \frac{3}{2}\frac{(1+K)k_1^2\exp(\chi_1+\chi_2)+k_1^2\exp(\chi_1)(1+K\exp(2\chi_2))}{[1+\exp(\chi_1)+\exp(\chi_2)+K\exp(\chi_1+\chi_2)]^2}, \tag{84}$$

which is though bounded does not fall off to zero as $\chi_2 \to \infty$. One can proceed further [29] and obtain an $(1,N)$ dromion of the form

$$v_{(1,N)} = \frac{3}{2}\frac{(K-1)k_1\exp(\chi_1)\sum_{i=2}^{N+1} l_i\exp(\chi_i)}{[1+\sum_{j=1}^{N}\exp(\chi_j)+K\exp(\chi_1)\sum_{i=2}^{N+1}\exp(\chi_i)]^2}. \tag{85}$$

Finally, as in the case of the generalized $(2+1)$ dimensional KdV equation (§ 5.4), one can obtain more general exponentially localized solution of the form

$$u = \tfrac{3}{8}k_1^2\,\mathrm{sech}^2\tfrac{1}{2}[k_1x+g(\xi)], \tag{86a}$$

$$v = \tfrac{3}{8}k_1h(\xi)\,\mathrm{sech}^2\tfrac{1}{2}[k_1x+g(\xi)], \tag{86b}$$

where $g(\xi)$ and $h(\xi)$ are arbitrary functions in the variable $\xi = y - k_1^2t$. Choosing $h(\xi)$ suitably, one can indeed obtain a very large class of localized structures for the potential v.

Thus in the case of the breaking soliton equation (77), it is not the physical field which is localized but it is rather the potential which admits localized solutions. Looking from another point of view, one may take (78) along with (77b) as the basic evolution equation, in which case we have now the new physical field which admits localized structures.

6.3 $(2+1)$ *dimensional NLS equation*

Considering the symmetric generalized $(2+1)$ dimensional nonlinear Schrödinger equation (29) introduced by Fokas [17], it can be rewritten as

$$iq_t + (\alpha+\beta)q_{\eta\eta} - (\alpha-\beta)q_{\xi\xi} - 2\lambda q[(\alpha+\beta)V-(\alpha-\beta)U] = 0, \tag{87a}$$

$$|q|^2_\eta = V_\xi, \tag{87b}$$

$$|q|^2_\xi = U_\eta. \tag{87c}$$

Equation (87) contains three important systems:

(i) $\alpha = \beta = \frac{1}{2}$: Simplest complex scalar equation in $(2+1)$ dimensions
(ii) $\alpha = 0,\ \beta = 1$: Davey–Stewartson equation I (DSI)
(iii) $\alpha = 1,\ \beta = 0$: Davey–Stewartson equation III (DSIII).

$$V = -\lambda\partial_{\eta\eta}\log\phi, \quad U = -\lambda\partial_{\xi\xi}\log\phi, \tag{88b}$$

one obtains the bilinearized form

$$[iD_t - (\alpha-\beta)D_\xi^2 + (\alpha+\beta)D_\eta^2]g\cdot\phi = 0, \tag{89a}$$

$$D_\xi D_\eta\phi\cdot\phi = -2\lambda gg^*. \tag{89b}$$

Expanding

$$g = \varepsilon g_1 + \varepsilon^3 g_3 + \cdots, \quad \phi = 1 + \varepsilon^2\phi_2 + \varepsilon^4\phi_4 + \cdots \tag{90}$$

and following the usual procedure, the basic line solitons are obtained as

$$q(\xi,\eta,t) = \tfrac{1}{2}\exp(-\psi)\,\mathrm{sech}(\chi_{1R}+\psi)\exp(i\chi_{1I}), \tag{91a}$$

$$V(\xi,\eta,t) = -\lambda s_{1R}^2\mathrm{sech}^2(\chi_{1R}+\psi), \tag{91b}$$

$$U(\xi,\eta,t) = -\lambda p_{1R}^2\mathrm{sech}^2(\chi_{1R}+\psi), \tag{91c}$$

$$\chi_1 = p_1\xi + s_1\eta - i[(\alpha-\beta)p_1^2 - (\alpha+\beta)s_1^2]t + \chi_1^{(0)}. \tag{91d}$$

One identifies two nonparallel ghost solitons, one at $\eta = -\infty$,

$$U = -\lambda p_{1R}^2\mathrm{sech}^2[\bar{\chi}_{1R}+\psi], \quad \bar{\chi}_{1R} = p_{1R}[\xi + 2(\alpha-\beta)p_{1I}t], \tag{92a}$$

and the other at $\xi = -\infty$.

$$V = -\lambda s_{1R}^2\mathrm{sech}^2[\hat{\chi}_{1R}+\psi'], \quad \hat{\chi}_{1R} = s_{1R}[\eta + 2(\alpha+\beta)s_{1I}t]. \tag{92b}$$

Consequently,the exponentially localized (1,1) dromion is obtained with the choice

$$\phi_{11D} = 1 + j\exp(\chi_1+\chi_1^*) + k\exp(\chi_2+\chi_2^*) + l\exp(\chi_1+\chi_1^*+\chi_2+\chi_2^*), \tag{93a}$$

with

$$\chi_1 = p_1\xi - i(\alpha-\beta)p_1^2t + c_1, \quad \chi_2 = s_1\eta + i(\alpha+\beta)s_1^2t + c_2, \tag{93b}$$

where p_1, s_1, c_1 and c_2 are constants. Choosing g appropriately, one then obtains

$$q_{11D} = \frac{\rho\exp(p_{1R}\hat{\xi} + s_{1R}\hat{\eta} + i\{p_{1I}\hat{\xi} + s_{1I}\hat{\eta} + [(\alpha+\beta)|s_1|^2 - (\alpha-\beta)|p_1|^2]\})}{1 + j\exp(2p_{1R}\hat{\xi}) + k\exp(2s_{1R}\hat{\eta}) + l\exp(2[p_{1R}\hat{\xi} + s_{1R}\hat{\eta}])}, \tag{94a}$$

where

$$\hat{\xi} = \xi + 2(\alpha-\beta)p_{1I}t, \quad \hat{\eta} = \eta - 2(\alpha+\beta)s_{1I}t. \tag{94b}$$

The multidromions can also be constructed by generalizing the above procedure. Similarly exponentially localized breather solutions also follow by putting $s_{1I} = p_{1I} = 0$ in (94a) to obtain

$$q_{11B} = \frac{\rho \exp(p_{1R}\xi + s_{1R}\eta + i\{(\alpha+\beta)s_{1R}^2 - (\alpha-\beta)p_{1R}^2\}t)}{1 + j\exp(2p_{1R}\xi) + k\exp(2s_{1R}\eta) + l\exp(2[p_{1R}\xi + s_{1R}\eta])}. \tag{95}$$

6.4 *The* $(2+1)$ *dimensional generalized NLS equation used by Strachan [18]*

The system of equations (30) is completely integrable and satisfies the Painlevé property [27]. It can be bilinearized through the transformation

$$q = \frac{g}{\phi}, \quad q^* = \frac{g^*}{\phi}, \quad v = 2\partial_{xy}\log\phi \tag{96}$$

as

$$iD_t g\cdot\phi = D_x D_y g\cdot\phi, \tag{97a}$$

$$D_x^2\phi\cdot\phi = 2gg^*. \tag{97b}$$

While admitting line solitons, it does not admit two nonparallel ghost solitons driving the boundaries in the usual sense. Thus the basic dromion solutions of the form considered in the previous examples do not seem to arise in this case.

However, one finds that a more general form of localized solutions corresponding to curved solitons can exist. To generate such localized solutions, we expand g and ϕ in the form of a power series similar to (90) to give rise to the following set of equations:

$$\varepsilon : ig_t^{(1)} = g_{xy}^{(1)} \tag{98a}$$

$$\varepsilon^2 : \phi_{xx}^{(2)} = g^{(1)}g^{(1)*} \tag{98b}$$

and so on. Solving (98a), we obtain

$$g^{(1)} = \sum_{j=1}^{N}\exp(\chi_j), \quad \chi_j = k_j x + m_j(y,t) + c_j, \tag{99a}$$

where $m_j(y,t)$ is an arbitrary function of (y,t) chosen such that

$$m_j(y,t) = m_j(\rho_j) = m_j(y - ik_j t) \tag{99b}$$

and k_j and c_j are complex constants. To construct one soliton solution, we take $N=1$ and substitute $g^{(1)}$ in (98b) to give

$$\phi^{(1)} = \exp(\chi_1 + \chi_1^* + 2\psi), \quad \exp(2\psi) = \frac{1}{4k_R^2}. \tag{100}$$

Hence, the physical field q and the potential V assume the form

$$q = k_{1R}\,\mathrm{sech}(\chi_{1R} + \psi)\exp(i\chi_{1I}), \tag{101a}$$

$$V = 2k_{1R}(m_{1R})_{\rho_{1R}}\,\mathrm{sech}(\chi_{1R} + \psi), \quad \rho_{1R} = y + k_{1I}t, \quad k_{1I} = Im(k_1). \tag{101b}$$

It is evident from (101) that both the physical field q and the potential V remain finite on the curve

$$C = \chi_{1R} + \psi = k_{1R}x + m_{1R}(y,t) + c_{1R} = 0 \tag{102}$$

and decay exponentially everywhere (as $x, y \to \infty$) apart from the curve $C = 0$ given by (102). We call such line solitons as "curved solitons". As $(m_{1R})_{\rho_{1R}}$ is arbitrary, one can choose it conveniently as

$$(m_{1R})_{\rho_{1R}} = \operatorname{sech}^2(\rho_{1R}) \tag{103}$$

so that a one dromion solution for the potential V becomes

$$V = 2k_{1R}\operatorname{sech}^2(\rho_{1R})\operatorname{sech}^2(\chi_{1R} + \psi). \tag{104}$$

One calls such localized solutions as "induced dromions" as they are induced by virtue of the arbitrary function present in the system. One can indeed construct a large class of localized solutions for the potential by choosing the arbitrary function properly.

6.5 $(2+1)$ *dimensional sine-Gordon equation*

The system (31) under the transformations

$$q = -\frac{i\sigma}{2}\theta_\eta, \quad r = -\frac{i\sigma}{2}\theta_\xi \tag{105}$$

can be recast in the form

$$q_{\xi t} + \tfrac{1}{2}\rho_\xi q + \tfrac{1}{2}\rho_\eta r = 0, \tag{106a}$$

$$r_{\xi t} + \tfrac{1}{2}\rho_\xi q + \tfrac{1}{2}\rho_\eta r = 0, \tag{106b}$$

$$\sigma^2\rho_{\xi\eta} = -2(qr)_t. \tag{106c}$$

Again the P-property can be established algorithmically [30]. One then obtains the bilinearizing transformation

$$q = \frac{q}{\phi}, \quad r = \frac{h}{\phi}, \tag{107a}$$

$$\rho = 2\partial_t(\log\phi) + 2\int_{-\infty}^{\xi} m_1(\xi',t)\,\mathrm{d}\xi' + 2\int_{-\infty}^{\eta} m_1(\eta',t)\,\mathrm{d}\eta' \tag{107b}$$

so that the Hirota bilinear form of eq. (106) becomes

$$D_\xi D_t g\cdot\phi + m_2(\xi,t)g\cdot\phi + m_1(\eta,t)h\cdot\phi = 0, \tag{108a}$$

$$D_\eta D_t g\cdot\phi + m_2(\xi,t)g\cdot\phi + m_1(\eta,t)h\cdot\phi = 0, \tag{108b}$$

$$\sigma^2 D_\xi D_\eta\phi\cdot\phi = -2gh, \tag{108c}$$

$$hD_\eta D_t\phi\cdot\phi - gD_\xi D_t\phi\cdot\phi = 0, \tag{108d}$$

where $m_1(\eta,t)$ and $m_2(\xi,t)$ are certain arbitrary functions which can be identified with

Considering the 2DSGI ($\sigma^2 = +1$) case, we note that the presence of nonzero m_1 and m_2 are essential for the formation of line kink solitons as well as localized solutions. Expanding g, h and ϕ in a suitable formal power series, one obtains the line kinks for the case in which m_1 and m_2 are non-zero constants as

$$q = \frac{2iq_1 \exp\chi_1}{1+\exp(2\chi_1 + 2\delta)}, \quad r = \frac{2ip_1 \exp\chi_1}{1+\exp(2\chi_1+2\delta)}, \tag{109}$$

where p_1, q_1 and δ are constants and

$$\chi_1 = 2p_1\xi + 2q_1\eta - \frac{m_1}{2q_1}t - \frac{m_2}{2p_1} + \chi_1^{(0)}, \quad \chi_1^{(0)}\text{: consts.} \tag{110}$$

Then we have

$$\theta(\xi,\eta,t) = 4\tan^{-1}[\exp(\chi_1+\delta)]. \tag{111}$$

Taking m_1 and m_2 in the form

$$m_1(\eta,t) = m_1^0 + m_1'(\eta'), \quad \eta' = \eta + v_1 t \tag{112a}$$

$$m_2(\xi,t) = m_2^0 + m_1'(\xi'), \quad \xi' = \xi + v_2 t \tag{112b}$$

one ultimately obtains the $(1,1)$ dromion solution [30, 37]

$$\theta = 4\tan^{-1}\left\{\frac{\zeta\exp[-(p_1\xi' + q_1\eta')]}{(1+\exp[-2(p_1\xi'-\delta_1)])(1+\exp[-2(q_1\eta'-\delta_2)])}\right\}. \tag{113}$$

6.6 *The* $(2+1)$ *dimensional long dispersive wave (2DLW) equation*

Considering the 2DLDW equation (32), we find that again the *P*-property is true and that with the transformation

$$q = \frac{g}{\phi}, \quad r = \frac{h}{\phi}, \quad v = \frac{\partial^2}{\partial x^2}\log\phi \tag{114}$$

it can be bilinearized in the form

$$(\lambda D_t + D_x^2) g\cdot\phi = 0, \tag{115a}$$

$$(\lambda D_t - D_x^2) h\cdot\phi = 0, \tag{115b}$$

$$D_x D_\eta \phi\cdot\phi = -2gh. \tag{115c}$$

Analysis of (115) leads to the line soliton solutions of the form

$$q = \frac{\exp(-p_1 x + s_1\eta - (p_1^2/\lambda)t + c_1)}{1+\exp(-[p_1+p_1']x + [s_1+s_1']\eta + (1/\lambda)[p_1'^2 - p_1^2]t + c + 2\psi)}, \tag{116a}$$

$$r = \frac{\exp(-p_1' x + s_1'\eta + (p_1'^2/\lambda)t + c_1')}{1+\exp(-[p_1+p_1']x + [s_1+s_1']\eta + (1/\lambda)[p_1'^2 - p_1^2]t + c + 2\psi)}, \tag{116b}$$

while the potential v is described by the line soliton

$$v = \frac{-(p_1 + p_1')^2}{4}\, \mathrm{sech}^2 \tfrac{1}{2}(\chi_1 + \chi_1' + 2\psi), \tag{116c}$$

where

$$\chi_1 = -p_1 x + s_1 \eta - \frac{p_1^2}{\lambda} t + c_1, \quad \chi_1' = -p_1' x + s_1' \eta + \frac{p_1'^2}{\lambda} t + c_1',$$

$$\exp(2\psi) = \frac{1}{(p_1 + p_1')(s_1 + s_1')}. \tag{116d}$$

We also note that the composite field is given by

$$Q = qr = \frac{(p_1 + p_1')(s_1 + s_1')}{4}\, \mathrm{sech}\, \tfrac{1}{2}(\chi_1 + \chi_1' + 2\psi). \tag{117}$$

Further in the limit $(s_1 + s_1') \to 0$ (or equivalently at the boundary $\eta = -\infty$), one has the ghost soliton

$$v = \frac{(p_1 + p_1')^2}{4}\, \mathrm{sech}^2 \frac{1}{2}\left([p_1 + p_1']x + \frac{1}{\lambda}(p_1'^2 - p_1^2)t + c_1 \right). \tag{118}$$

The above features indicate that it is not the physical fields q and r that will get localized, but it is rather the composite field $Q =$ "qr" which will get localized. Indeed with the choice

$$\phi = 1 + \exp \chi_1 + \exp \chi_2 + K \exp(\chi_1 + \chi_2), \tag{119a}$$

$$\chi_1 = p_1 x - \frac{p_1^2 t}{2} + c_1, \quad \chi_2 = s_1 \eta + c_2, \tag{119b}$$

one obtains the (1,1) dromion for the composite field as

$$Q = \text{“}qr\text{”} = \int_{-\infty}^{x} v_\eta \, \mathrm{d}x' = \frac{(1 - K) p_1 s_1 \exp(\chi_1 + \chi_2)}{[1 + \exp(\chi_1) + \exp(\chi_2) + K \exp(\chi_1 + \chi_2)]^2}, \quad K > 0. \tag{120}$$

More general $(1, N)$ dromions can be given by

$$(qr)_{1N} = \frac{(1 - K) p_1 \exp(\chi_1) \sum_{i=1}^{N} s_i \exp(\chi_{i+1})}{[1 + \sum_{j=1}^{N+1} \exp(\chi_j) + K \exp\left(\chi_1 \sum_{i=1}^{N} \exp(\chi_{i+1})\right)]^2}. \tag{121}$$

Again, it should be mentioned that one can indeed generate induced localized solutions for the composite field by utilising the arbitrary function $f(\eta)$ present in the system.

7. Conclusions

In this paper, we have investigated the type of basic nonlinear coherent structures that can arise in a class of nonlinear evolution equations in $(2 + 1)$ dimensions. In contrast to the ordinary solitons in $(1 + 1)$ dimensions, it was pointed out that much richer structures can arise in $(2+1)$ dimensions: line solitons, lump solitons, exponentially localized dromions and their generalizations and so on. In particular, we have discussed in detail certain algorithmic methods of obtaining the localized dromion-like solutions, either in the field variables or in the potential variables or even in the composite variables. In all these cases, the relevance of boundary contributions to form localized structures was stressed.

The necessity of boundary contributions to form exponentially localized solutions also justifies the type of NLEEs we have discussed here, namely all of them have nonlocal terms or equivalently extra fields associated with them. Simple minded $(2+1)$ dimensional (local) generalizations of $(1+1)$ dimensional soliton equations do not seem to admit exact analytical localized structures in general. More work needs to be carried out to understand the basic structures in such systems and the analysis of such systems will constitute an important area of development in nonlinear dynamics in future.

Acknowledgements

This work forms part of a research project sponsored by the Department of Science and Technology, Government of India.

References

[1] N J Zabusky and M D Kruskal, *Phys. Rev. Lett.* **15**, 240 (1965)
[2] C S Gardner, J M Greene, M D Kruskal and R M Miura, *Phys. Rev. Lett.* **19**, 1095 (1967)
[3] M J Ablowitz and P A Clarkson, *Solitons, Nonlinear Evolution Equations and Inverse Scattering* (Cambridge University Press, Cambridge, 1991)
[4] M Lakshmanan (ed.), *Solitons: Introduction and Applications*, (Springer-Verlag, Berlin, 1988)
[5] M Lakshmanan (ed.) *Chaos, Solitons and Fractals* (Special issue on Solitons in Science and Engineering: Theory and Applications) **5**, 2213-2656 (1995)
[6] P L Bhatnagar, *Nonlinear Waves in One-Dimensional Dispersive Systems* (Oxford University Press, Calcutta, 1979)
[7] M Boiti, J J P Leon, L Martina and F Pempinelli, *Phys. Lett.* **A132**, 432 (1988)
[8] A S Fokas and P M Santini, *Physica* **D44**, 99 (1990)
[9] M J Ablowitz and A S Fokas, *Stud. Appl. Math.* **69**, 135 (1983)
[10] A S Fokas and M J Ablowitz, *Stud. Appl. Math.* **69**, 211 (1983)
[11] A S Fokas and M J Ablowitz, *J. Math. Phys.* **25**, 2494 (1984)
[12] B G Konopelchenko, *Solitons in Multidimensions* (Springer-Verlag, Berlin, 1993)
[13] V E Zakharov and S V Manakov, *Funct. Anal. Appl.* **19**, 89 (1985)
[14] F Calogero, *Lett. Nuovo Cimento* **14**, 443 (1975)
[15] M Boiti, J J P Leon, M Manna and F Pempinelli, *Inv. Prob.* **2**, 271 (1986)
[16] S P Novikov and A P Veselov, *Physica* **D18**, 267 (1986)
[17] A S Fokas, *Inv. Prob.* **10**, L19 (1994)
[18] I A B Strachan, *Inv. Prob.* **8**, L21 (1992); *J. Math. Phys.* **34**, 243 (1993)
[19] B G Konopelchenko and C Rogers, *Phys. Lett.* **A158**, 391 (1991); *J. Math. Phys.* **34**, 214 (1993)
[20] J J C Nimmo, *Phys. Lett.* **A168**, 113 (1992)
[21] S Chakravarty, S L Kent and E T Newman, *J. Math. Phys.* **36**, 763 (1995)
[22] M Lakshmanan and R Sahadevan, *Phys. Rep.* **224**, 1 (1993)
[23] M J Ablowitz, A Ramani and H Segur, *J. Math. Phys.* **21**, 715 (1980)
[24] M Lakshmanan, *Int. J. Bifurcation and Chaos* **3**, 3 (1993)

[26] M Daniel, M D Kruskal, M Lakshmanan and K Nakamura, *J. Math. Phys.* **33**, 771 (1992)
[27] R Radha and M Lakshmanan, *Inv. Problems* **10**, L29 (1994)
[28] R Radha and M Lakshmanan, *J. Math. Phys.* **35**, 4746 (1994)
[29] R Radha and M Lakshmanan, *Phys. Lett.* **A197**, 7 (1995)
[30] R Radha and M Lakshmanan, *J. Phys.* **A29**, 1551 (1996)
[31] R Radha amd M Lakshmanan, *Chaos, Solitons and Fractals* **8**, 17 (1997)
[32] R Radha and M Lakshmanan, *J. Math. Phys.* **38**, 292 (1997)
[33] R Radha and M Lakshmanan, *J. Phys.* **A30**, (1997) to appear
[34] R Radha, Ph.D Thesis (Bharathidasan University, 1996)
[35] Sen-yue Lou, *J. Phys.* **A28**, 7227 (1995)
[36] J Hietarinta, *Phys. Lett.* **A149**, 113 (1990)
[37] V Dubrovsky and B G Konopelchenko, *Inv. Prob.* **9**, 391 (1993)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 1,
January 1997
pp. 189–204

Geometry and nonlinear evolution equations

RADHA BALAKRISHNAN
The Institute of Mathematical Sciences, Madras 600 113, India

Abstract. We briefly review the nonlinear dynamics of diverse physical systems which can be described in terms of moving curves and surfaces. The interesting connections that exist between the underlying differential geometry of these systems and the corresponding nonlinear partial differential equations are highlighted by considering classic examples such as the motion of a vortex filament in a fluid and the dynamics of a spin chain.The association of the dynamics of a non-stretching curve with a hierarchy of completely integrable soliton-supporting equations is discussed. The application of the surface embeddability approach is shown to be useful in obtaining such connections as well as exact solutions of some nonlinear systems such as the Belavin–Polyakov equation and the inhomogeneous Heisenberg chain.

Keywords. Nonlinear dynamics; geometry.

PACS Nos **03.40; 75.10; 02.40**

1. Introduction

The connection between differential geometry and certain nonlinear partial differential equations (NLPDE) has attracted the attention of both physicists and mathematicians for over a century now. Nineteenth century geometers had shown such a relationship between the structure of minimal surfaces (i.e., surfaces whose mean curvature is zero) and the hyperbolic Liouville equation. Another classic example is the connection [1] between the geometry of pseudospherical surfaces (i.e., surfaces with negative constant Gaussian curvature) and the sine-Gordon equation. This result achieved a special significance many years later, when it was realized that these equations were integrable by the powerful inverse scattering transform (IST) method discovered in 1967 by Gardner *et al* [2]. This observation inspired a lot of interesting work on the relationship [3] between the differential geometry of special surfaces and integrable soliton-supporting equations [4]. Around the same time, the differential geometry of moving curves [3] and their relation to integrable equations was also studied. Interestingly, this latter connection first arose in the context of a *physical* application in the field of hydrodynamics [5]. Pioneering work by Hasimoto [5] showed that the equation of motion of a vortex filament regarded as a space curve was equivalent to the well-known, integrable nonlinear Schrödinger equation (NLSE) [4]. Using the Hasimoto transformation that relates space curves and complex curvature functions, Lamb [6] generalized the above result by demonstrating the link between the motion of certain space curves and soliton-bearing equations. Soon afterwards, Lakshmanan [7] showed that the Landau–Lifshitz equation which describes the continuum dynamics of the classical ferromagnetic Heisenberg chain was gauge-equivalent to the NLSE .

In recent years, there has been renewed interest in looking for such connections. It has been shown that the dynamics of a non-stretching planar curve can be connected with the integrable, modified Korteweg–de Vries (MKdV) hierarchy [8]. Another example is the Belavin–Polyakov equation [9] satisfied by a unit vector field, which appears in a variety of physical applications such as the nonlinear sigma model, the two-dimensional Heisenberg ferromagnet, etc. Its equivalence to the exactly solvable elliptic Liouville equation has been established [10]. A surface embeddability approach has been demonstrated to be useful in finding explicit solutions to the spin evolution equation of an inhomogeneous Heisenberg chain [11]. In this paper, we briefly review some of the above-mentioned examples of nonlinear dynamical systems which can be described geometrically. The purpose is to illustrate how the application of the classical differential geometry of curves and surfaces can be used to find such surprising connections as well as exact dynamical solutions.

2. Space curves and surfaces

We begin by summarizing some of the salient features of the space-curve formalism and surface theory, which will be useful in the applications to be considered in subsequent sections.

2.1 *Time evolution of twisted space curves*

A *static* space curve in E^3 is described either by its parametric equations or by its natural equations: $K = K(s)$ and $\tau = \tau(s)$, where K, τ and s are the curvature, the torsion and the arc length of the space curve. Consider a curve which in its parametric form is described by the position vector $\mathbf{x} = \mathbf{x}(s)$. Let $\mathbf{t}(s) = \mathrm{d}\mathbf{x}/\mathrm{d}s$ be the unit tangent to this curve, and $\mathbf{n}$ and $\mathbf{b}$ the unit normal and binormal, so that $\mathbf{t}, \mathbf{n}$ and $\mathbf{b}$ form an orthogonal triad on the curve. They satisfy the well known Frenet–Serret equations [1]

$$\mathbf{t}_s = K\mathbf{n}, \quad \mathbf{n}_s = -K\mathbf{t} + \tau\mathbf{b}, \quad \mathbf{b}_s = -\tau\mathbf{n} \tag{2.1}$$

where the subscript denotes $(\mathrm{d}/\mathrm{d}s)$ and K and τ are given by

$$K^2 = \mathbf{t}_s \cdot \mathbf{t}_s, \quad \tau = \mathbf{t} \cdot (\mathbf{t}_s \times \mathbf{t}_{ss})/K^2. \tag{2.2}$$

Let this curve *evolve* in time t in such a way that the local 'velocity' at s is specified by

$$\mathbf{V} = (\partial\mathbf{x}(s,t)/\partial t) = \mathbf{x}_t = \alpha\mathbf{t} + \beta\mathbf{n} + \gamma\mathbf{b}. \tag{2.3}$$

The scalar functions α, β and γ represent the tangential, normal and binormal components of $\mathbf{V}$. As we shall see, their functional form will be determined by the physics of the model under consideration.

There is another aspect of curve motion. As the tangent to a static curve is $\mathbf{x}_s = \mathbf{t}$ by definition, it is clear that in any physical application involving moving curves, it is possible to write down the time evolution of $\mathbf{t}$ from (2.3) by using

$$\mathbf{t}_t = \mathbf{x}_{st}. \tag{2.4}$$

In some other models (like spin chains), $\mathbf{t}_t$ can be written down directly. In general, since $\mathbf{t}$ is a unit vector,

$$\mathbf{t}_t = g\mathbf{n} + h\mathbf{b}, \tag{2.5}$$

where g and h will be local functions of the curvature K and the torsion τ. Further, when the compatibility conditions

$$\mathbf{t}_{ts} = \mathbf{t}_{st}, \quad \mathbf{n}_{ts} = \mathbf{n}_{st}, \quad \mathbf{b}_{ts} = \mathbf{b}_{st} \tag{2.6}$$

are imposed, one typically obtains *coupled nonlinear partial differential equations for K and τ*. In many applications, these turn out to be well-known exactly solvable equations. In such cases, the analysis shows that the above property arises essentially due to certain special geometrical features with which the curve evolution is endowed.

2.2 *Surface theory*

Consider a surface generated by a position vector $\mathbf{r}(u,v)$ which is a function of two parameters (local coordinates) u and v. The metric, i.e., the first fundamental form, is given in the usual notation [1] as $I = (\mathrm{d}\mathbf{r})^2 = E(\mathrm{d}u)^2 + 2F\mathrm{d}u\,\mathrm{d}v + G(\mathrm{d}v)^2$, where $E = (\mathbf{r}_u)^2$, $F = \mathbf{r}_u\cdot\mathbf{r}_v$ and $G = (\mathbf{r}_v)^2$. The unit normal to the surface is defined by $\mathbf{n} = (\mathbf{r}_u \times \mathbf{r}_v)/|\mathbf{r}_u \times \mathbf{r}_v|$. The extrinsic curvature tensor, i.e., the second fundamental form is defined as $II = -\mathrm{d}\mathbf{r}\cdot\mathrm{d}\mathbf{n} = L(\mathrm{d}u)^2 + 2M\mathrm{d}u\,\mathrm{d}v + N(\mathrm{d}v)^2$, where $L = \mathbf{n}\cdot\mathbf{r}_{uu}, M = \mathbf{n}\cdot\mathbf{r}_{uv}$ and $N = \mathbf{n}\cdot\mathbf{r}_{vv}$. At every point on the surface, we can introduce a moving trihedron made up of the three linearly independent vectors $\mathbf{r}_\mathrm{u}$, $\mathbf{r}_\mathrm{v}$ and $\mathbf{n}$. Since every vector can be linearly expressed in terms of these three basis vectors, one obtains the following Gauss–Weingarten (GW) equations [1]:

$$\mathbf{r}_{uu} = \Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{11}\mathbf{r}_v + L\mathbf{n} \tag{2.7a}$$

$$\mathbf{r}_{uv} = \Gamma^1_{11}\mathbf{r}_u + \Gamma^2_{12}\mathbf{r}_v + M\mathbf{n} \tag{2.7b}$$

$$\mathbf{r}_{vv} = \Gamma^1_{22}\mathbf{r}_u + \Gamma^2_{22}\mathbf{r}_v + N\mathbf{n} \tag{2.7c}$$

$$\mathbf{n}_u = p_1\mathbf{r}_u + p_2\mathbf{r}_v \tag{2.7d}$$

$$\mathbf{n}_v = q_1\mathbf{r}_u + q_2\mathbf{r}_v \tag{2.7e}$$

where the expressions for the Christoffel symbols Γ^k_{ij} and the coefficients p_i, q_i, $i = 1,2$ are given below for ready reference

$$\begin{aligned}
&\Gamma^1_{11} = [GE_u - 2FF_u + FE_v]/2\Lambda; \quad \Gamma^1_{12} = [GE_v - FG_u]/2\Lambda \\
&\Gamma^1_{22} = [2GF_v - GG_u - FG_v]/2\Lambda; \quad \Gamma^2_{11} = [2EF_u - EE_v - FE_u]/2\Lambda \\
&\Gamma^2_{12} = [EG_u - FG_v]/2\Lambda; \quad \Gamma^2_{22} = [EG_v - 2FF_v + FG_u]/2\Lambda \\
&p_1 = [MF - LG]/\Lambda; \quad p_2 = [LF - ME]/\Lambda. \\
&q_1 = [NF - MG]/\Lambda; \quad q_2 = [MF - NE]/\Lambda,
\end{aligned} \tag{2.8}$$

where $\Lambda = (EG - F^2)$. The Gauss–Mainardi–Codazzi (GMC) equations [1] are obtained from the compatibility conditions $\mathbf{r}_{uuv} = \mathbf{r}_{uvu}$ and $\mathbf{r}_{uvv} = \mathbf{r}_{vvu}$. We will write down the relevant equations when we discuss specific applications. The fundamental theorem of surfaces states that if the surface parameters E, F, G, L, M and N satisfy the GMC equations, then the corresponding surface $\mathbf{r}(u,v)$ is uniquely determined except for its position in space. We will see how this can be used in applications where the GMC equations turn out to be exactly solvable equations, with the surface parameters determined exactly as explicit functions of u and v.

3. Vortex filament motion and the Hasimoto transformation

It has been observed that vortex filaments in a fluid very often preserve their identity when they move. In this section, we summarize Hasimoto's work [5] which shows that on using a local induction approximation (to be explained below), the propagation of a soliton along a filament can be obtained.

A thin vortex filament in a fluid can be regarded as a space curve. The fluid is incompressible and inviscid and has a non-vanishing vorticity

$$\mathbf{w}(\mathbf{x}) = \nabla \times \mathbf{V}(\mathbf{x}), \tag{3.1}$$

where $\mathbf{V}(\mathbf{x})$ is the fluid velocity at a point with position vector $\mathbf{x}$ on the filament. Equation (3.1) implies the following Biot–Savart law:

$$\mathbf{V}(\mathbf{x}) = \int \mathrm{d}^3x' \mathbf{w}(\mathbf{x}') \times (\mathbf{x}-\mathbf{x}')/|\mathbf{x}-\mathbf{x}'|^3. \tag{3.2}$$

Thus the filament is transported by the velocity *induced* by its own vorticity. For simplicity, it is assumed that $\mathbf{w}(\mathbf{x})$ is tangential to the filament at every point. Since the vortex has a thin core, the volume integral in (3.2) can be replaced by a line integral as follows:

$$\mathbf{V}_I(\mathbf{x}) \sim \int \mathbf{t}(\mathbf{x}') \times (\mathbf{x}-\mathbf{x}')/|\mathbf{x}-\mathbf{x}'|^3 \mathrm{d}s', \tag{3.3}$$

where $\mathbf{V}_I$ is the induced velocity and $\mathbf{x} = \mathbf{x}(s,t)$ is the position vector of a point on the filament curve at time t. Because of the thinness of the core, the contribution of the local portion of the filament dominates the Biot–Savart integral. Expanding $\mathbf{x}'$ in a Taylor series around $\mathbf{x}(s,t)$,

$$\mathbf{x}'(s',t) = \mathbf{x}(s,t) + (s'-s)\mathbf{x}_s + (1/2)(s'-s)^2\mathbf{x}_{ss} + O(s'-s)^3. \tag{3.4}$$

Since $\mathbf{x}_s = \mathbf{t}$ and $\mathbf{x}_{ss} = \mathbf{t}_s = K\mathbf{n}$ from (2.1), substituting (3.4) in (3.3) yields

$$\mathbf{V}_I \sim K(\mathbf{t} \times \mathbf{n}) = K\mathbf{b}, \tag{3.5}$$

Comparing (3.5) with (2.3), we see that in this 'local induction approximation', the velocity $\mathbf{V}_I = \mathbf{x}_t$ of the vortex filament has no tangential and normal components but only a component along the binormal, proportional to the curvature K. Equation (3.5) leads to

$$\mathbf{t}_t = K_s\mathbf{b} - K\tau\mathbf{n}, \tag{3.6}$$

$$K_t = -2K_s\tau - K\tau_s \tag{3.7a}$$

$$\tau_t = [(K_{ss}/K) - \tau^2]_s - KK_s. \tag{3.7b}$$

On using the Hasimoto transformation

$$\psi = K \exp\left\{ i \int^s \tau \mathrm{d}s \right\}, \tag{3.8}$$

Equations (3.7) reduce to the nonlinear Schrödinger equation (NLSE)

$$i\psi_t + \psi_{ss} + (1/2)|\psi|^2\psi = 0 \tag{3.9}$$

whose exact solutions are known [4]. The NLSE is a completely integrable evolution equation, associated with an infinite number of integrals of motion, strict N-soliton solutions, etc. From the soliton solution, the curvature and torsion of the filament can be read off using (3.8). Thus, the local induction approximation is seen to lead to a shape- and velocity-preserving motion of a helical vortex filament in the fluid. It is interesting that such a motion of vortex filaments has been experimentally observed in certain types of fluids. Finally, this example shows that the special geometry underlying the completely integrable NLSE is encoded in the simple geometric relationship $\mathbf{x}_t = K\mathbf{b}$ [Eq. (3.5)] for the local velocity of a space curve moving in E^3. More recently, it has been shown [12] that going beyond the local induction approximation can lead to a connection with more general soliton-bearing equations.

4. Dynamics of the classical Heisenberg ferromagnetic chain

The subject of quasi-one dimensional magnets with various symmetries is interesting because such systems can be fabricated in the laboratory. In this section we review the isotropic ferromagnetic chain.

Consider a one-dimensional lattice with classical spins $\mathbf{S}_i$ of constant magnitude (i.e., $(\mathbf{S}_i)^2 = S^2 = 1$) at every site i, with the nearest neighbour interaction Hamiltonian

$$H = -J \sum_i \mathbf{S}_i \cdot \mathbf{S}_{i+1}, \quad J > 0. \tag{4.1}$$

The spin evolution equation derived from (4.1) is $(\mathrm{d}\mathbf{S}_i/\mathrm{d}t) = J\mathbf{S}_i \times (\mathbf{S}_{i+1} + \mathbf{S}_{i-1})$. On taking the continuum limit $\mathbf{S}_i \to \mathbf{S}(x,t)$, this reduces to the nonlinear partial differential equation (the Landau–Lifshitz equation)

$$\mathbf{S}_t = J\mathbf{S} \times \mathbf{S}_{xx}, \tag{4.2}$$

where the subscripts denote partial derivatives. The underlying geometry of (4.2) is most conveniently analysed by identifying $\mathbf{S}$ with the unit tangent to a curve [13], so that the Landau–Lifshitz equation describes a *moving space curve*. (Note that the instantaneous spin configuration on the chain can be mapped to a corresponding space curve at that instant). Thus (4.2) can be cast in the form (after rescaling time t to Jt)

$$\mathbf{t}_t = \mathbf{t} \times \mathbf{t}_{ss}. \tag{4.3}$$

Pioneering work by Lakshmanan [7] showed that Eq. (4.3) is gauge-equivalent to the NLSE. The following interesting connection between the present interacting spin problem and the (seemingly unrelated) vortex filament motion considered in § 3 is worth noting: The filament evolution $\mathbf{x}_t = K\mathbf{b}$ (Eq. (3.5)) when differentiated with respect to s leads to (4.3)! Thus (4.3) also yields the *same* coupled equations (3.7) found in the fluid mechanical context, and thence to the NLSE. The soliton solutions play the role of nonlinear excitations in the spin chain [7]. It must be noted that the solution of NLSE contains information about the energy and current densities [13]. In this example we have seen that the geometry of the NLSE is encoded in the special time evolution of the *tangent* vector as given by (4.3).

Before concluding this section, we mention that there are some extensions of the Heisenberg chain which can be analysed through the space curve formalism-for instance, the inclusion [14] of the Landau–Gilbert damping term in the Hamiltonian. Integrability in higher dimensions for the spherically symmetric case has also been recently investigated [15].

5. Dynamics of nonstretching planar curves

There are many physical applications involving curve dynamics which preserve certain global geometric quantities such as the length of an open curve or the area enclosed by a closed curve. Examples are polymers, cell membranes or closed vortex filaments. An interesting paper by Goldstein and Petrich [8] considers the dynamics of curves on a *plane*. In this case the torsion $\tau = 0$, and Eqs (2.1) become

$$\mathbf{t}_s = K\mathbf{n}, \quad \mathbf{n}_s = -K\mathbf{t}. \tag{5.1}$$

Now

$$\mathbf{x}_t = U\mathbf{n} + W\mathbf{t}, \tag{5.2}$$

where U and W are the normal and tangential components of the velocity. If these are chosen to be local functions of the curvature K and its derivatives, i.e.,

$$U = U(K, K_s, K_{ss} \ldots) \quad \text{and} \quad W = W(K, K_s, K_{ss} \ldots),$$

we get from (5.2)

$$\mathbf{x}_{st} = \mathbf{t}_t = (W_s - KU)\mathbf{t} + (U_s + KW)\mathbf{n}.$$

Since $\mathbf{t}.\mathbf{t}_t = 0$, this leads to

$$W_s = KU, \quad W = \int KU\mathrm{d}s + C, \tag{5.3}$$

where $C =$ constant. Hence, in general,

$$\mathbf{t}_t = (U_s + KW)\mathbf{n}. \tag{5.4}$$

At this stage, one poses the following question: How should a planar curve evolve in time so as to keep its length $L = \int \mathrm{d}s$ fixed? One way to ascertain $L_t = 0$ is to ensure that $\mathrm{d}s/\mathrm{d}t = 0$ so that the curve does not 'stretch' as time evolves. This imposes the compatibility conditions $\mathbf{t}_{st} = \mathbf{t}_{ts}$ and $\mathbf{n}_{st} = \mathbf{n}_{ts}$. Using (2.1), (5.3) and (5.4) in these yields the following evolution equation for K:

$$K_t = U_{ss} + K^2 U + K_s \left\{ \int KU \mathrm{d}s + C \right\} = -\Omega U, \tag{5.5}$$

where the operator Ω is given by

$$\Omega = (-\partial_{ss} + K^2 + K_s \partial_s^{-1} K) \tag{5.6}$$

with

$$\partial_s^{-1} KU = \int KU \mathrm{d}s + C.$$

It is clear that the evolution of the curve is decided by the choice of U and W in (5.2). Since U and W are related by (5.3) in general, and W should be a local function, either U should vanish or KU must be a total derivative. The simplest choice $U = U^{(1)} = 0$ gives $W = W^{(1)} = C$ and

$$K_t = -CK_s. \tag{5.7}$$

This is a linear equation representing the trivial motion of the curve along itself with velocity C. (Note that the velocity $\mathbf{x}_t^{(1)} = C\mathbf{t}$ has only a tangential component). The next choice $U = U^{(2)} = -K_s$ gives $W = W^{(2)} = (\frac{1}{2}K^2) + C$, giving the modified Korteweg-de Vries (MKdV) equation

$$K_t = -K_{sss} + (3/2)K^2 K_s \tag{5.8}$$

for the choice $C = 0$. This is a completely integrable equation. The one-soliton solution $K(s,t) = \sqrt{u}$ sech $(\sqrt{u}(s - ut))$ represents a moving localized profile (lump) traveling along the curve with velocity $u > 0$. Note that the velocity $\mathbf{x}_t^{(2)} = -K_s \mathbf{n} - (\frac{1}{2}K^2)\mathbf{t}$ has both tangential *and* normal components.

Proceeding likewise, i.e., by choosing the right-hand side of each equation for K to be the normal velocity U of the succeeding equation, a *hierarchy* of equations is obtained, the next one being

$$K_t = -K_{sssss} - (15/8)K^4 K_s - (5/2)K_s^3 - (5/2)K^2 K_{ss} - 10KK_s K_{ss}, \tag{5.9}$$

for the choice $U = U^{(3)} = K_{sss} + \frac{3}{2}K^2 K_s$.

Note that the order of the equation increases at each stage. If the curve is closed, the area enclosed is $A = \frac{1}{2}\oint(\mathbf{x} \times \mathbf{x}_s)\mathrm{d}s$, giving $A_t = \oint U \mathrm{d}s$. Since U is a total derivative for all members of the hierarchy, the corresponding evolutions preserve not only the perimeter but also the area enclosed. In addition, due to complete integrability, the dynamics also possesses an infinite number of integrals of the motion which are polynomials in the curvature and their derivatives. (There could of course be other choices like $U = K^n K_s$, $W = -K^{n+2}/(n+2)$, which would also conserve these global quantities, but it is not known whether the resulting dynamics would be integrable). In conclusion, the recursion operator Ω defined in (5.6) has a universal geometrical significance in defining the curvature evolution under arclength-conserving dynamics.

It has also been pointed out [6, 16, 17] that the Frenet–Serret equations are equivalent to the Ablowitz–Kaup–Newell–Segur (AKNS) scattering problem in the IST with zero eigenvalue. We conclude this section by mentioning that Langer and Perline [18] have shown that the dynamics of a nonstretching filament regarded as a non-planar space curve can give rise to an analogous recursion operator for the NLSE hierarchy, on using the Hasimoto transformation.

6. The Belavin–Polyakov equation

The Belavin–Polyakov equation [9]

$$\mathbf{m}_y = \mathbf{m}_x \times \mathbf{m}, \quad (\mathbf{m})^2 = 1 \tag{6.1}$$

arises in several physical applications. In these coupled nonlinear partial differential equations, both the independent variables x and y could be spatial variables, or one spatial and the other temporal, depending on the physical context. The static solutions of the nonlinear sigma model in $(2+1)$ dimensions, the static two-dimensional Heisenberg ferromagnet, and the Heisenberg antiferromagnet in $(1+1)$ dimensions [19] are examples where the relevant configurations are classical *unit* vectors $\mathbf{m}$ satisfying (6.1).

Let us first apply the moving space curve formulation to this problem [10]. It is possible to identify $\mathbf{m}$ with the tangent vector to a space curve which 'evolves' either spatially or temporally, depending on whether y is a space variable or a time variable. Thus (6.1) can be written in the form

$$\mathbf{t}_u = \mathbf{t}_s \times \mathbf{t} = -K\mathbf{b}, \tag{6.2}$$

on using (2.1). Writing $\mathbf{n}_u = a_1\mathbf{t} + a_2\mathbf{b}$ and $\mathbf{b}_u = g_1\mathbf{t} + g_2\mathbf{n}$, the compatibility conditions (2.6) determine $a_1 = 0$, $a_2 = (-K_s/K)$, $g_1 = K$ and $g_2 = (K_s/K)$. Further, the coupled evolution equations for K and τ are

$$K_u = K\tau \tag{6.3a}$$

and

$$\tau_u = (-K_s/K)_s - K^2. \tag{6.3b}$$

These equations combine to give

$$(K_u/K)_u + (K_s/K)_s = -K^2, \tag{6.3c}$$

which becomes the elliptic Liouville equation

$$\phi_{ss} + \phi_{uu} = -\exp(2\phi), \tag{6.3d}$$

on setting $\phi = \ln K$. Since the exact general solution of (6.3d) is known [10], the exact solutions for the curvature $K = \exp\phi$ and $\tau = \phi_u$ of the evolving space curve can be determined. Thus the use of classical differential geometry of space curves has enabled us to establish [10] the exact solvability of (6.2), by connecting it to the exactly solvable elliptic Liouville equation.

This connection can also be found [10] using surface theory. By identifying the solution $\mathbf{m}(x,t)$ of (6.1) with a position vector $\mathbf{r}(u,v)$, it is possible to describe a surface. The relevant expressions for the first and second fundamental forms have been given in § 2. Using them in conjunction with eq. (6.1), we see that $F = 0$ and

$G = (\mathbf{m}_y)^2 = (\mathbf{m}_x \times \mathbf{m})^2 = E$. Thus the metric has a *conformal* ("*isothermal*") form. The unit normal to the surface $\mathbf{n}$ is defined as in § 2. From (6.1), we obtain $(\mathbf{m}_x \times \mathbf{m}_y) = -E\mathbf{m}$. Hence $|\mathbf{m}_x \times \mathbf{m}_y| = E$. A short calculation shows that the second fundamental form parameters are given by $L = N = E$ and $M = 0$. Thus the Gaussian curvature $K = (LN - M^2)/(EG - F^2)$ and the mean curvature $H = \frac{1}{2}(EN + GL - 2FM)/(EG - F^2)$ are both constant. With $\mathbf{m}(x, y)$ as the position vector creating the surface, the Gauss–Weingarten equations (Eqs (2.7)) become

$$
\begin{aligned}
\mathbf{m}_{xx} &= (E_x/2E)\mathbf{m}_x - (E_y/2E)\mathbf{m}_y - E\mathbf{m}\\
\mathbf{m}_{xy} &= (E_y/2E)\mathbf{m}_x + (E_x/2E)\mathbf{m}_y\\
\mathbf{m}_{yy} &= -(E_x/2E)\mathbf{m}_x + (E_y/2E)\mathbf{m}_y - E\mathbf{m}\\
\mathbf{n}_x &= -\mathbf{m}_x\\
\mathbf{n}_y &= -\mathbf{m}_y
\end{aligned}
\tag{6.4}
$$

We now obtain the Gauss–Mainardi–Codazzi equations [1] from the various compatibility conditions as mentioned in § 2. Note that $\mathbf{n}_{xy} = \mathbf{n}_{yx}$ is automatically satisfied, while the conditions $\mathbf{m}_{xxy} = \mathbf{m}_{xyx}$ and $\mathbf{m}_{xyy} = \mathbf{m}_{yyx}$ both reduce, after a short calculation, to the following equation for the coefficient E of the conformal metric:

$$(E_y/2E)_y + (E_x/2E)_x = -E. \tag{6.5}$$

Defining $\phi = \frac{1}{2}\ln E$, the above equation becomes the elliptic Liouville equation (6.3d). Its *general* solution can be written down as

$$\phi(x, y) = \tfrac{1}{2}\ln\{-4A'(z^*)B'(z)/[A(z^*) + B(z)]^2\} \tag{6.6}$$

where $z = x + iy, z^* = x - iy, A'(z^*) = \mathrm{d}A/\mathrm{d}z^*$, $B'(z) = \mathrm{d}B/\mathrm{d}z$ and A, B are arbitrary functions. Hence

$$E(x, y) = -4A'(z^*)B'(z)/[A(z^*) + B(z)]^2. \tag{6.7}$$

The fundamental theorem of surfaces states that if the metric and curvature tensors satisfy the Gauss–Mainardi–Codazzi equations (which in the present case is the elliptic Liouville equation), then the Gauss–Weingarten equations can in principle be integrated to determine the surface (specified by $\mathbf{m}(x, y)$) uniquely, up to a rotation and translation in space. Therefore (6.1) is exactly solvable. Some interesting special solutions have been found in Ref. [19].

It can be shown that the basic equations (6.3a) and (6.3b) determined using the 'direct' moving curve method can be combined into the following equation for the complex Hasimoto function $q = \kappa \exp\left(i\int_{-\infty}^{s} \tau \mathrm{d}s\right)$:

$$iq_u - q_s - q\int_{-\infty}^{s} |q|^2 \mathrm{d}s = 0. \tag{6.8}$$

This equation had in fact been obtained earlier [19] in another context. A short calculation verifies that its imaginary part yields (6.3a), while its real part gives (6.3b). These equations when combined become the elliptic Liouville equation (6.3d), as we have already shown above.

7. The inhomogeneous Heisenberg chain

The examples discussed in the previous sections show how classical differential geometry can be used as an effective tool to find the connection of a nonlinear system described by a vector evolution equation with a completely integrable NLPDE or with one whose exact general solution is known. As mentioned in the introduction, while these connections are indeed useful in predicting similar properties for the corresponding vector evolution equation, finding the explicit solution of the latter is not always an easy task. In many physical situations, finding special exact solutions explicitly would be of interest (the corresponding nonlinear system need not necessarily be always completely integrable). In this section we consider the example of the inhomogeneous Heisenberg chain to show how exact solutions can be found using an appropriate geometric approach [11].

Several years ago, this author proposed a model Hamiltonian [20]

$$H = -\sum_{i} f_i\, \mathbf{S}_i\cdot\mathbf{S}_{i+1},\, (\mathbf{S}_i)^2 = S^2 = 1, \tag{7.1}$$

where $\mathbf{S}_i$ denotes the spin vector and f_i the site dependent (inhomogeneous) exchange interaction between the spins. The continuum evolution for $\mathbf{S}$ was found to be

$$\mathbf{S}_t = (f\mathbf{S}\times\mathbf{S}_x)_x, \quad \mathbf{S}^2 = 1, \tag{7.2}$$

By associating $\mathbf{S}$ with the tangent to a moving curve, it was shown that the above equation is gauge-equivalent to a generalized nonlinear Schrödinger equation (GNLSE) of the form

$$iq_t + (fq)_{xx} + 2q\left\{f|q|^2 + \int_{-\infty}^{x} f_x|q|^2 \mathrm{d}x\right\} = 0. \tag{7.3}$$

Here q is given essentially by the Hasimoto function defined in (3.8), namely,

$$q = (\kappa/2)\exp\int_{-\infty}^{x} \tau\, \mathrm{d}x, \tag{7.4}$$

with the curvature κ and the torsion τ of the curve being given by

$$\kappa^2 = (\mathbf{S}_x)^2, \quad \tau = \mathbf{S}\cdot(\mathbf{S}_x\times\mathbf{S}_{xx})/\kappa^2 \tag{7.5}$$

respectively. There are only two forms of $f(x)$ for which the conventional IST method is directly applicable to the above GNLSE, *viz.*, (i) when f is a constant (in which case it reduces to the usual NLSE), and (ii) when f is a linear function [21] of x. All other $f(x)$ fall into a special class [20]. In a interesting recent paper, Cieśliński *et al* [22] have used some ideas from Sym's surface approach [3] to find exact solutions of (7.2) and (7.3) for several other forms of the inhomogeneity function $f(x)$. Representing $\mathbf{S}$ as the spatial derivative of a vector $\mathbf{r}$, i.e., setting

$$\mathbf{S} = \mathbf{r}_x, \tag{7.6}$$

it can be verified that the kinematic equation

$$\mathbf{r}_t = f(\mathbf{r}_x\times\mathbf{r}_{xx}), \quad \mathbf{r}_x^2 = 1. \tag{7.7}$$

implies Eq. (7.2). Assuming the surface generated by the position vector $\mathbf{r}(x,t)$ to be equipped with a metric of the geodesic form, it was shown [22] that the Gauss–Mainardi–Codazzi (GMC) equations [1] for this surface can be cast in the same form as (7.3), upon making appropriate identifications between q and f on the one hand, and the coefficients of the first and second fundamental forms (i.e., the metric and the extrinsic curvature, respectively) on the other. By specializing to surfaces of revolution, it was demonstrated that the equations for the geodesics on the surface could be integrated explicitly for certain cases, and an algorithm was found for obtaining both f and the corresponding solution $\mathbf{S}$. When the geodesic coordinates were taken to be x and t, the solutions called 'spins-on-meridians' could be found for a wide variety of time-independent, bounded, positive functions $f(x)$, i.e., inhomogeneous ferromagnetic couplings.

In what follows we use [11] a geometric formulation different from the above approach in the sense that, although we too identify $\mathbf{r}(x,t)$ with a position vector that generates a surface, we do not compare the GMC equations to (7.3), nor do we use the general equations of geodesics. Our strategy is to regard the basic kinematic equation (7.7) for $\mathbf{r}$ as a *constraint* on the surface generated. Such a constraint is expected to permit only certain special geometries for the surface [3]. Indeed, for the model under investigation, the surface metric is shown to be *necessarily* of the geodesic form, and further, certain coefficients of the two fundamental forms get related through the function f. Using these results in the GMC equations, it becomes possible to 'integrate' the latter and find the expressions for L, M and N (the coefficients of the second fundamental form) as well as the corresponding function f in terms of the metric coefficient G, its x-derivatives and two (arbitrary) integration constants, for time-independent metrics. It is then demonstrated that the solution q of (7.3) can also be written in terms of the above-mentioned quantities, simply by re-expressing the moving curve parameters κ and τ in terms of surface coefficients by using the Gauss-Weingarten (GW) equations [1] for the surface. Thus we see that given an arbitrary metric coefficient, the explicit solution for q along with the corresponding f can be written down. This solution q is a *complex* function in general, and, as already mentioned in § 4, is interesting in its own right since it contains information [11] on the energy and current densities along the chain. Furthermore, on inspecting the expressions for L, M and N obtained from the GMC equations, it is readily seen that they correspond to those of surfaces of revolution, [1] when one of the two integration constants referred to above vanishes. Also, $G^{1/2}$ plays the role of the generator of revolution, and the explicit solution of $\mathbf{r}(x,t)$ can be written down in this limiting case. In this limit, our expression for $f(x)$ reduces to the result obtained [22] in the geodesic approach, for the 'spins-on-meridians' solutions. The corresponding solutions for q are necessarily *real*. For certain common surfaces of revolution like the sphere, the torus and the catenoid, the solution $\mathbf{S}$ of Eq. (7.2) can be expressed solely as a functional of the inhomogeneity function f that appears in the equation, essentially because the surface metric can also be written down as a functional of f for these surfaces. We now present a few details of this analysis:

We start with the kinematic equation for $\mathbf{r}(x,t)$ Eq. (7.7), and identify $\mathbf{r}(x,t)$ with a position vector generating a smooth surface in E^3, with local coordinates x and t. Note that $\mathbf{S}^2 = \mathbf{r}_x^2 = 1$ in the model. From the evolution, it is readily verified that $\mathbf{r}_x . \mathbf{r}_t = 0$. We use the same notation (E,F,G,L,M,N) for the metric and the extrinsic curvature parameters, respectively, as in § 6. It is easy to verify that the metric is *necessarily* constrained to be of the geodesic form given below:

$$I = (\mathrm{d}x)^2 + G(x,t)(\mathrm{d}t)^2. \tag{7.8}$$

The unit normal **n** to the surface is given by

$$\mathbf{n} = (\mathbf{r}_x \times \mathbf{r}_t)/G^{1/2}. \tag{7.9}$$

As already mentioned in § 2, the GW equations for a surface can be written in terms of L, M, N and the usual Christoffel symbols Γ^k_{ij}. For the metric given in (7.8), the latter reduce to (see Eqs (2.8)):

$$\left.\begin{aligned} &\Gamma^1_{11} = \Gamma^2_{11} = \Gamma^1_{12} = 0; \\ &\Gamma^2_{12} = \tfrac{1}{2}G_x/G; \quad \Gamma^1_{22} = -\tfrac{1}{2}G_x \quad \text{and} \quad \Gamma^2_{22} = \tfrac{1}{2}G_t/G. \end{aligned}\right\} \tag{7.10}$$

Using these, the GW equations (equations (2.7)) read

$$\mathbf{r}_{xx} = L\mathbf{n} \tag{7.11a}$$

$$\mathbf{r}_{xt} = \tfrac{1}{2}(G_x/G)\mathbf{r}_t + M\mathbf{n} \tag{7.11b}$$

$$\mathbf{r}_{tt} = -\tfrac{1}{2}G_x\mathbf{r}_x + \tfrac{1}{2}(G_t/G)\mathbf{r}_t + N\mathbf{n} \tag{7.11c}$$

$$\mathbf{n}_x = -L\mathbf{r}_x - (M/G)\mathbf{r}_t \tag{7.11d}$$

$$\mathbf{n}_t = -M\mathbf{r}_x - (N/G)\mathbf{r}_t. \tag{7.11e}$$

Further, for this (geodesic) surface it can be shown that

$$(\mathbf{r}_x \times \mathbf{r}_{xx}) = L(\mathbf{r}_x \times \mathbf{n}) = (L/G^{1/2})\mathbf{r}_t. \tag{7.12}$$

Comparing this with the given equation (7.7), we get the constraint

$$f = -G^{1/2}/L, \tag{7.13}$$

for $L \neq 0$. In other words: for a given f, the evolution can have a solution only if the surface coefficients G and L are related through f as above.

For a surface with $E = 1$ and $F = 0$, the compatibility conditions $(\mathbf{r}_{xx})_t = (\mathbf{r}_{xt})_x$ and $(\mathbf{r}_{tt})_x = (\mathbf{r}_{xt})_t$ yield the following GMC equations:

$$\begin{aligned} -(LN - M^2)/G = (\Gamma^2_{12})_x - (\Gamma^2_{11})_t + \Gamma^1_{12}\Gamma^2_{11} - \Gamma^1_{11}\Gamma^2_{12} \\ + \Gamma^2_{12}\Gamma^2_{12} - \Gamma^2_{11}\Gamma^2_{22} \end{aligned} \tag{7.14a}$$

$$L_t - M_x = L\Gamma^1_{12} + M(\Gamma^2_{12} - \Gamma^1_{11}) - N\Gamma^2_{11} \tag{7.14b}$$

$$M_t - N_x = L\Gamma^1_{22} + M(\Gamma^2_{22} - \Gamma^1_{12}) - N\Gamma^2_{12} \tag{7.14c}$$

On using the expressions for Γ^k_{ij} given in (7.10), and the definition

$$K = (LN - M^2)/G, \tag{7.15}$$

of the Gaussian curvature K, Eqs (7.14) reduce to

$$-K = (G_x/2G)_x + (G_x/2G)^2 \tag{7.16a}$$

$$L_t - M_x = M(G_x/2G) \tag{7.16b}$$

$$M_t - N_x = -L(G_x/2) + M(G_t/2G) - N(G_x/2G). \tag{7.16c}$$

$f = G^{1/2}/L$ is satisfied. We note that the Gauss equation can be written in the more convenient form,

$$(G^{1/2})_{xx} = -KG^{1/2}. \tag{7.17}$$

Details of the determination of G, L, M, N are given elsewhere [11]. The equations above can be analysed to yield

$$f = \pm\phi\ (2A_o - \phi_x^2 - C_o\phi^{-2})^{1/2}/(C_o\phi^{-3} - \phi_{xx}) \tag{7.18}$$

$$L = -\phi/f = (\phi_{xx} - C_o\phi^{-3})/(2A_o - \phi_x^2 - C_o\phi^{-2})^{1/2} \tag{7.19}$$

$$M = C_o\phi^{-1} \tag{7.20}$$

$$N = f(\phi_{xx} - C_o\phi^{-3}) = -\phi(2A_o - \phi_x^2 - C_o\phi^{-2})^{1/2}, \tag{7.21}$$

where

$$\phi = G^{1/2}. \tag{7.22}$$

These expressions satisfy the GMC equations, of course, since they have essentially been obtained by integrating the latter equations. The point to be noted is that the quantities L, M and N as well as f, depend on ϕ and its derivatives. Thus, *given* an arbitrary function ϕ (with the proviso that $(2A_o - \phi_x^2 - C_o\phi^{-2}) > 0$), we can use (7.18) to (7.21) to find these quantities explicitly. Combining this result with the fundamental theorem of surfaces, we conclude that for inhomogeneity functions f determined from (7.18) in this fashion, exact solutions $\mathbf{r}(x,t)$ of (7.7) exist. In addition, the solution q corresponding to such functions can be found explicitly. Explicit solutions for $\mathbf{r}$ can also be found for a special sub-class of f, as we shall see.

As mentioned in the beginning, the solution q of (7.3) can be expressed in terms of the curvature κ and torsion τ, the parameters of a (moving) curve, as in (7.4). Using these definitions together with the GW equations (7.11), we get

$$\kappa = (\mathbf{S}_x \cdot \mathbf{S}_x)^{1/2} = (\mathbf{r}_{xx} \cdot \mathbf{r}_{xx}) = L. \tag{7.23}$$

τ can also be expressed in terms of the surface coefficients after some algebra. We obtain

$$\tau = \mathbf{S} \cdot (\mathbf{S}_x \times \mathbf{S}_{xx})/\kappa^2 = \mathbf{r}_x \cdot (\mathbf{r}_{xx} \times \mathbf{r}_{xxx})/L^2 = M/G^{1/2}. \tag{7.24}$$

Substituting these expressions in the Hasimoto relation (7.4) gives the following solution for q in terms of L, M and G:

$$q = \tfrac{1}{2} L \exp i \int MG^{-1/2} \mathrm{d}x. \tag{7.25}$$

We have already shown that L and M are given by (7.19) and (7.20), for $G = G(x)$. Substituting these quantities in (7.25) and using (7.22) we get

$$q = \tfrac{1}{2}(\phi_{xx} - C_o\phi^{-3})(2A_o - \phi_x^2 - C_o\phi^{-2})^{-1/2} \exp\left[iC_o \int \phi^{-2}\mathrm{d}\right]. \tag{7.26}$$

Note that the torsion $\tau = C_o\phi^{-2}$ for this system. Summarizing, we have the following result: For any arbitrary ϕ (such that $(2A_o - \phi_x^2 - C_o\phi^{-2}) > 0$), the inhomogeneity function f can be found from Eq. (4.11), and for that f, the corresponding solution q is given by eq. (7.26).

Next, we discuss solutions for **r** in the two possible classes of surfaces corresponding, respectively, to $M = 0$ and $M \neq 0$:

(i) $M = 0$:

It is evident from (7.20) that C_o must vanish when M vanishes. Using this, (7.19) and (7.21) yield

$$L = \phi_{xx}/(2A_o - \phi_x^2)^{1/2} \quad \text{and} \quad N = -\phi(2A_o - \phi_x^2)^{1/2}. \tag{7.27}$$

It is verified easily that these extrinsic curvature coefficients correspond to the following *surface of revolution* [1]:

$$\mathbf{r} = \left(\int (2A_o - \phi_x^2)^{1/2}\mathrm{d}x,\ \phi\cos t,\ \phi\sin t\right). \tag{7.28}$$

Thus ϕ plays the role of the generator of a surface of revolution. Further, since $\mathbf{r}_x^2 = \mathbf{S} = 1$, (7.28) leads to $2A_o = 1$. By computing $\mathbf{r}_{xx}$ from it we find $L^2 = (\mathbf{r}_{xx}\cdot\mathbf{r}_{xx}) = 2A_o\phi_{xx}^2/(2A_o - \phi_x^2)$, which agrees with the expression for L given in (7.27), on setting $2A_o = 1$. (The expression for N can be verified similarly). Using $A_o = \frac{1}{2}$, $C_o = 0$, in Eq. (7.18), we get

$$f = \pm\phi(1 - \phi_x^2)^{1/2}/\phi_{xx}. \tag{7.29}$$

This expression for f was also obtained by Cieśliński *et al* [22], by integrating the equations for the geodesics on a surface of revolution. We have thus established that their result corresponds to the limit $M = C_o = 0$ of the more general expression for f given in Eq. (7.18), which we derived as the inhomogeneity function supporting an exact dynamical solution for a time-dependent metric. For $C_o = 0$, Eq. (7.26) reduces to $q = \frac{1}{2}\phi_{xx}/(2A_0 - \phi_x^2)^{1/2}$.

(ii) $M \neq 0$:

This corresponds to $C_0 \neq 0$ (see Eq. (7.20)). Since surfaces of revolutions necessarily have $M = 0$, a non-vanishing M represents some other class of surfaces. The construction of **r** in accord with (7.14)–(7.21) with $C_o \neq 0$ is non-trivial and is an open problem. However, the solution for q can be found from Eq. (7.26).

Finally, we present the solutions for **S** in terms of the Gaussian curvature K. Equation (7.17) for K can be written as

$$K = -\phi_{xx}/\phi. \tag{7.30}$$

Using this in (7.29), we obtain

$$Kf = (1 - \phi_x^2)^{1/2}. \tag{7.31}$$

Substituting (7.31) in the expression for **r** in (7.28) and computing $\mathbf{r}_x$ yields, with

$2A_o = 1,$

$$\begin{aligned}\mathbf{S} = \mathbf{r}_x &= ((1-\phi_x^2)^{1/2}, \phi_x \cos t, \phi_x \sin t)\\ &= (Kf, (1-K^2f^2)^{1/2}\cos t, (1-K^2f^2)^{1/2}\sin t).\end{aligned} \tag{7.32}$$

Hence **S** is given as a functional of the product Kf. For the cases of a torus, a sphere and a catenoid, the corresponding solutions have been found explicitly in Ref. 11.

8. Concluding remarks

Understanding the relationship between nonlinear analysis and geometry is currently an active field of research. Many interesting results have not been considered here due to limitations of space. For instance, it has been shown [23] recently that if a non-stretching curve moves on an N-dimensional sphere in such a way that its dynamics does not depend explicitly on the radius of the sphere, the MKdV hierarchy, the NLSE hierarchy and their multi-component generalizations arise in a natural fashion, for $N = 2, N = 3$ and $N > 3$ respectively. New connections have been established using the differential geometry of constant mean curvature surfaces [24]. Another aspect which we have not discussed in this paper is that of a geometric phase associated with moving curves [19] and solitons [25]. Interesting parallels have been found between elastic dynamics of ribbons [26] and the appearance of such anholonomic phases.

Our emphasis in this paper has been on showing how physical applications can be instructive in suggesting ways of obtaining interesting mathematical connections between geometry and nonlinearity. The seminal work of Hasimoto [5] in fluid dynamics and its subsequent novel developments [27] represent a case in point.The added advantage of analysing a physical system is the possibility of observing such exact mathematical solutions (e.g., solitons) in the laboratory [28]. It is clear that the underlying nonlinearity in the system plays a crucial role also in deciding the topological character of the solutions. For example, the kink configurations of the sine-Gordon equation, which are topologically non-trivial, get replaced by (topologically) uninteresting plane wave solutions when the equation is linearized. We close with the remark that it would be of great interest to understand the connection between nonlinearity and differential geometry using physical applications in higher dimensions.

References

[1] D J Struik, *Lectures on classical differential geometry*, 2nd Edition (Dover, New York, 1961)
M Spivak, *A comprehensive introduction to differential geometry*, (Publish or Perish, Boston, 1970) Vols. 1–4
[2] C S Gardner, J M Greene, M D Kruskal and R M Miura, *Phys. Rev. Lett.* **19**, 1095 (1967)
[3] F Lund and T Regge, *Phys. Rev.* **D14**, 1524 (1976)
G L Lamb, *Phys. Rev. Lett.* **37**, 235 (1976)
M Lakshmanan, *J. Math. Phys.* **20**, 1667 (1979)
M Lakshmanan and K M Tamizhmani, *Appl. Sci. Res.* **37**, 127 (1981) and references therein;
A. Sym, *Lett. Nuovo Cimeto.* **33**, 394 (1982)
[4] See for instance, M J Ablowitz and H Segur, *Solitons and the inverse scattering transform* (SIAM, Philadelphia, 1981)

[5] H Hasimoto, *J. Fluid Mech.* **51**, 477 (1972)
[6] G L Lamb, *J. Math. Phys.* **12**, 1548 (1977)
[7] M Lakshmanan, *Phys. Lett.* **A61**, 53 (1977)
[8] R E Goldstein and D M Petrich, *Phys. Rev. Lett.* **67**, 3203 (1991)
[9] A Belavin and A M Polyakov, *JETP Lett.* **22**, 245 (1975)
[10] Radha Balakrishnan, *Phys. Lett.* **A204**, 243 (1995)
[11] Radha Balakrishnan and P Guha, *J. Math. Phys.* **37**, 3651 (1996)
[12] Y Fukumoto and T Miyazaki, *J. Fluid Mech.* **159**, 369 (1991)
[13] M Lakshmanan, Th W Ruijgrok and C J Thomson, *Physica* **A84**, 577 (1976)
[14] M Daniel and M Lakshmanan, *Physica* **A120**, 125 (1983)
[15] M Daniel, K Porsezian and M Lakshmanan, *J. Math. Phys.* **35**, 6438 (1994)
[16] M Lakshmanan, *Phys. Lett.* **A64**, 353 (1978); *J. Math. Phys.* **20**, 1667 (1979)
M Lakshmanan and S Ganesan, *Physica* **A132**, 117 (1985)
[17] K Nakayama, H Segur and M Wadati, *Phys. Rev. Lett.* **69**, 260 (1992)
[18] J Langer and R Perline, *J. Nonlinear Sci.* **1**, 71 (1991)
[19] Radha Balakrishnan, A R Bishop and R Dandoloff, *Phys. Rev. Lett.* **64**, 2107 (1990); *Phys. Rev.* **B47**, 3108 (1993)
[20] Radha Balakrishnan, *J. Phys.* **C15**, L1305 (1982) *Phys. Lett.* **A92**, 243 (1982), **E122**, 494 (1987); *Physica* **D16**, 405 (1985)
[21] M Lakshmanan and R K Bullough, *Phys. Lett.* **A80**, 287 (1980)
[22] J Cieslinski, A Sym and W Wesselius, *J. Phys.* **26**, 1353 (1993)
[23] A Doliwa and P M Santini, *Phys. Lett.* **A185**, 373 (1994)
[24] A I Bobenko, *Math. Ann.* **290**, 209 (1991)
[25] Radha Balakrishnan, *Phys. Lett.* **A180**, 239 (1993)
[26] I Klapper and M Tabor, *J. Phys.* **A27**, 4919 (1994)
[27] R L Ricca, *Chaos* **3**, 83 (1993)
[28] J Bernasconi and T Schneider, *Physics in one dimension* (Springer Verlag, Berlin, 1981)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 1,
January 1997
pp. 205–229

Nonlinear chiral models: Soliton solutions and spatio-temporal chaos

M S SRIRAM and J SEGAR
Department of Theoretical Physics, University of Madras, Guindy Campus, Madras 600 025, India

Abstract. Nonlinear effective Lagrangian models with a chiral symmetry have been used to describe strong interactions at low energy, for a long time. The Skyrme model and the chiral quark-meson model are two such models, which have soliton solutions which can be identified with the baryons. We describe the various kinds of soliton states in these nonlinear models and discuss their physical significance and uses in this review. We also study these models from the view point of classical nonlinar dynamical systems. We consider fluctuations around the $B = 1$ soliton solutions of these models (B, being the baryon number) and solve the spherically symmetric, time-dependent systems. Numerical studies indicate that the phase space around the Skyrme soliton solution exhibits spatio-temporal chaos. It is remarkable that topological solitons signifying stability/order and spatio-temporal chaos coexist in this model. In contrast with this, the soliton of the quark-meson model is stable even for large perturbations.

Keywords. Nonlinear; chiral models; Skyrme; quarks; mesons; soliton; baryon; spatio-temporal chaos.

PACS Nos 12.39; 05.45; 12.39

1. Introduction

Most of the field theories of physical interest are highly nonlinear and though perturbative techniques are very successful in elementary particle physics, condensed matter physics and several other areas, one has to grapple with the nonlinearities in the underlying field theories, sooner or later. This is especially true of theories which have no small parameter in which a perturbative expansion can be made, like the models for strong interaction or strongly correlated electron systems. So one has to rely on non-perturbative methods for these. It has been found that some of these field theories have a rich structure at the classical level, thanks to their nonlinear nature. They possess localised, stable solutions with finite energy at the classical level. These are the soliton solutions in the field theory parlance. Examples are the kink solutions of the Sine–Gordon theory, the vortex solutions of the Abelian–Higgs model or Chern–Simons–Higgs model in $2 + 1$ dimensions of relevance to superconductivity, and the 't Hooft–Polyakov monopole solution of a $O(3)$ gauge theory (or any gauge theory based on a semisimple group in which $O(3)$ can be embedded), of relevance to proton decay and cosmological models based on grand unified theories [1]. All these soliton solutions have one thing in common; they have a topological number or a winding number associated with them and an associated conservation law. This accounts partly for their stability also.

There are some models with soliton solution which have nothing to do with any topology; the SLAC bag model [2] and the Friedberg–Lee model [3] are examples of important theories with non-topological soliton solutions.

Long before quantum chromodynamics (QCD) emerged as the theory of strong interactions, it had been recognised that strong interactions are approximately invariant under chiral symmetry. In fact, it was known that for describing hadronic physics, the $SU(2) \times SU(2)$ symmetry based models are better than the $SU(3)$ symmetry based models. In 1962, Skyrme proposed a Lagrangian model with purely pionic degrees of freedom, invariant under $SU(2) \times SU(2)$, and noted that the soliton solutions of this theory fall into different homotopy classes characterised by a winding number n [4]. Skyrme identified the winding number with the baryon number. It was a very bold and novel idea at that time, to describe fermions from a purely bosonic field theory. Later it was established that the Skyrme solitons should indeed be quantized as fermions for odd number of colour charges [5]. In general, one uses a collective quantization approach to quantize the classical soliton solutions of the model [6]. The properties of the nucleon deduced in this model are in reasonable agreement with the experimental values. The Skyrme model also has a classical solution with winding number 2, with an energy distribution characteristic of a diffuse torus [7,8]. When quantized, the lowest state has the quantum numbers of the deuteron, with the correct size, though the binding energy is far too large. These toroidal solutions exist for winding number 3, 4,... also [7]. Apart from these, other states are obtained when the Skyrme model is extended to $SU(N), N \geq 3$ [9, 10]. The Skyrme model has also been applied to meson-baryon scattering [11] and inter-nucleon potential [12] with some success. We describe the salient features of the Skyrme model in § 2.

Though the Skyrmion picture of baryons is very attractive, it has some problems both at the phenomenological as well as at the conceptual level. On the basis of general arguments grounded firmly in the non-perturbative structure of QCD, it appears that a purely topological soliton picture of nucleon may not be correct. The favoured structure for the baryon involves three valance quarks interacting chiral-invariantly with meson fields [13]. This leads us to the chiral quark-meson model (CQM) [14]. In this model also, there is a soliton solution with three valance quarks. Here the baryon number of the soliton is provided by the quarks explicitly. Though there are topological properties associated with the meson fields in this model, the soliton solution here is not topological in the same sense as in the Skyrme model. Phenomenologically, CQM is somewhat better than the Skyrme model. There are some close parallels between the Skyrme model and CQM, in that many novel soliton states of the former are present in the latter also. For instance, CQM also has a stable state with the quantum numbers of the deuteron and states with higher baryon number with toroidal energy distributions [15], just as in the Skyrme model. We review this model in § 3.

Gauge theories have been studied from the view point of nonlinear dynamical systems, in recent times. One considers spatially homogeneous solutions of the Yang–Mills (YM) theory and obtains a Hamiltonian system (with a finite number of degrees of freedom), which is then shown to be stochastic [16]. With the inclusion of Higgs fields, Chern–Simons term etc., gauge theories remain chaotic in the case of spatially first to study the fluctuations around the Wu–Yang monopole solution of the SU(2) YM theory. By considering a harmonic expansion of the fluctuation, the problem essentially reduces to the Fermi–Pasta–Ulam approach for coupled nonlinear oscillators [21]. It was shown that the phase space near the Wu–Yang static solution is a stochastic region implying that the spherically symmetric time dependent YM system is non-integrable. Joy and Sabir have shown that YM and the Yang–Mills–Higgs (YMH) theories are non-integrable by using the Painlevé analysis and also by a computation of the maximal Lyapunov exponent [22]. Systematic and detailed numerical work by Kawabe and Ohta showed that the classical YM theory is always chaotic and that there is no phase transition [23]. However, when the Higgs fields are introduced and the phase space around the 't Hooft–Polyakov monopole solution of the YMH theory considered, there is a phase transition when the perturbation parameter is varied [24].

It would be of interest to find whether chaos in the underlying theory of strong interactions, namely QCD, is revealed at the level of the effective Lagrangian models based on it, also. The stability property of the Skyrme soliton has been discussed in the literature [25]. When the fluctuations around the soliton solutions are small, they can be treated in a perturbative manner and satisfy a linearized set of equations. Skyrmions are topological objects and it is not surprising that these studies confirm their stability under small perturbations. This is what makes it a meaningful object, fit for quantization. Though large fluctuations may not be relevant for the static properties of an individual isolated Skyrmion, they are expected to play a role when they are subjected to external fields or when we consider interactions among them. This is our motivation for studying the fluctuations in a non-perturbative manner by solving the full nonlinear equations for the system, though at a classical level.

In § 4 we consider fluctuations around the $B = 1$ soliton solution of the Skyrme model and determine their time evolution numerically [26]. We also find the time variations of the amplitudes of the various harmonic modes of the fluctuations. Our initial fluctuations correspond to excitation of only one mode. We consider both fluctuations whose magnitudes are small compared to those of the static solitons, as well as large initial fluctuations which are comparable to the magnitude of the static soliton solution. In the case of the Skyrme model, for small initial fluctuations, the amplitude of the initial mode has a regular behaviour, whereas the amplitudes of the other modes increase slowly, on the average. The "radius" of the soliton oscillates around the static value. However, when the perturbation is large, the amplitude of the initial mode decreases gradually and other modes are excited substantially. The radius of the soliton also tends to increase with time on the average. All these indicate chaos in the system. So, both topological solitons (symbolizing a high degree of symmetry and regularity) and stochasticity coexist in the model.

We have studied the stability aspects of the chiral quark-meson model [26] in § 5. In contrast to the Skyrme model, we find that the soliton of the chiral quark-meson model is very stable even for large perturbations. The amplitudes of the initial modes remain constant for a long time and there is very little excitation of the other modes. The system remarks.

2. Baryons as topological solitons in the Skyrme model

The Lagrangian of the Skyrme model including the pion-mass is given by [27]

$$L_{\text{Skyrme}} = \frac{F_\pi^2}{4}\,\mathrm{Tr}\,[\partial_\mu U \partial^\mu U^\dagger] + \frac{1}{32e^2}\,\mathrm{Tr}\,[\partial_\mu U U^\dagger, \partial_\nu U^\dagger]^2 \\ + \frac{F_\pi^2 m_\pi^2}{4}\,\mathrm{Tr}\,[(U + U^\dagger) - 2]. \tag{1}$$

Here U is an $SU(2)$ matrix-valued function of the pionic field variables $\pi_a (a = 1, 2, 3)$ of the form

$$U = \exp\left(-i\frac{\boldsymbol{\tau}\cdot\boldsymbol{\pi}}{F_\pi}\right) \tag{2}$$

where $\tau_a (a = 1, 2, 3)$ are the Pauli matrices and F_π is the pion-decay constant, whose physical value is 93 MeV. e is a dimensionless coupling constant and m_π is the pion mass. Except for the last term, L_{Skyrme} is invariant under the chiral $SU_{\mathrm{L}}(2) \times SU_{\mathrm{R}}(2)$ transformations (left-handed and right-handed isospin transformations):

$$U \longrightarrow gUh^\dagger \tag{3}$$

where

$$g \in SU_{\mathrm{L}}(2) \quad \text{and} \quad h \in SU_{\mathrm{R}}(2). \tag{4}$$

As things stand, the Skyrme Lagrangian is one among a class of nonlinear σ-models describing the low energy interactions among pions, which are known to be invariant under $SU_{\mathrm{L}}(2) \times SU_{\mathrm{R}}(2)$ transformations.

Now, for finite energy configurations, $U \longrightarrow U_0$, a constant unitary matrix, as $r \longrightarrow \infty$. By a global rotation we can choose U_0 to be the identity matrix. Hence,

$$U \longrightarrow 1 \quad \text{as} \quad r \longrightarrow \infty. \tag{5}$$

Thus, we may think of all points at spatial infinity as being identified with a single point. Because of this, the configuration space R^3 of coordinates r becomes the three sphere S^3, on which U is well defined. Now as $U \in SU(2)$, it can be written as

$$U = n_o + i\boldsymbol{\tau}.\hat{n} \tag{6}$$

where,

$$n_o^2 + \mathbf{n}^2 = 1. \tag{7}$$

Hence the group $SU(2)$ is identified topologically with the three sphere S^3. Thus the fields U are functions on S^3 and takes values in S^3. Then all the finite energy configurations U fall into distinct homotopy classes of the map $S^3 \longrightarrow S^3$, characterized by a winding number n. The explicit expression for the winding number n is,

$$n = \frac{1}{24\pi^2}\,\epsilon_{ijk} \int \mathrm{d}^3x\,\mathrm{Tr}\,[U^\dagger \partial_i U U^\dagger \partial_j U U^\dagger \partial_k U]. \tag{8}$$

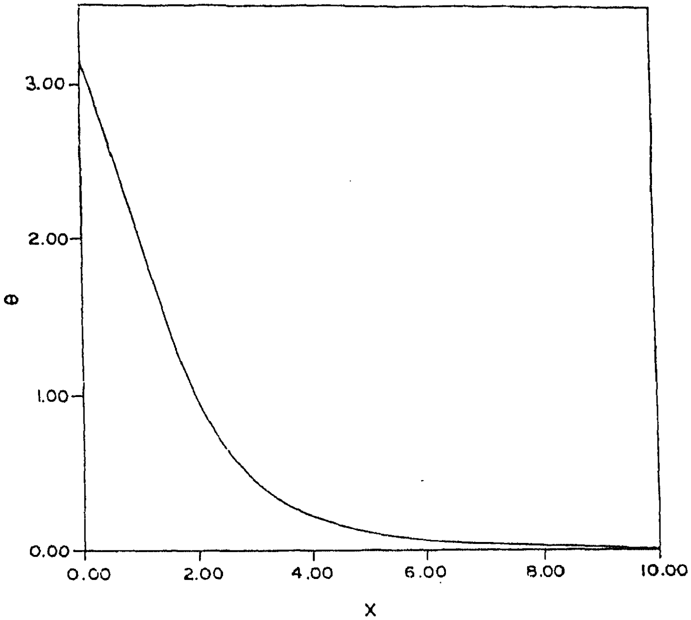


Figure 1. The Skyrme function $\Theta(x)$,corresponding to the hedgehog ansatz ($B = 1$).

Skyrme boldly conjectured that the static, localized, finite energy solutions or the solitons of the model correspond to the baryons and that the winding number n should be identified with the baryon number B. The second term in L_{Skyrme} (quartic term), known as the Skyrme term gives an elegant lower bound to the energy of the soliton and prevents it from collapsing to zero size. It is the stabilizing term.

The lowest energy soliton solution with $n = B = 1$, of the model is invariant under the combined spatial and isospin rotations:

$$-i(\mathbf{r}\times\nabla)_i U(\mathbf{r}) + \left[\frac{\tau_i}{2}, U(\mathbf{r})\right] = 0. \tag{9}$$

Then U has the form

$$U = \exp[i\tau\cdot\hat{r}\Theta(r)] = \cos\Theta(r) + i\tau\cdot\hat{r}\sin\Theta(r). \tag{10}$$

This is known as the spherically symmetric or 'hedgehog' ansatz. The winding number for this ansatz can be shown to be

$$n = B = \frac{1}{\pi}[\Theta(0) - \Theta(\infty)]. \tag{11}$$

Θ for the $B = 1$ soliton, to be identified with the nucleon satisfies, $\Theta(0) = \pi$ and $\Theta(\infty) = 0$. With the ansatz in equation (10), the equations of motion reduce to:

$$\left[\frac{x^2}{4}+2\sin^2\Theta\right]\Theta''+\frac{x\Theta'}{2}+\sin 2\Theta\,\Theta'^2-\frac{\sin 2\Theta}{4}-\frac{\sin^2\Theta\sin 2\Theta}{x^2}-\frac{x^2\beta}{4}\sin\Theta=0, \tag{12}$$

where we have introduced the dimensionless variables $x = 2eF_\pi r$ and $\beta = (m_\pi^2/4e^2F_\pi^2)$. Here, prime denotes differentiation with respect to x. The solution $\Theta(x)$ has the form given in figure 1. The size of the soliton solution is $\approx 1\,\mathrm{fm}$.

The Skyrme model Lagrangian has an extra discrete symmetry (apart from the chiral symmetry), which has no experimental basis. This additional symmetry is removed by adding the so called 'Wess–Zumino term' [5]. A careful treatment of the chiral anomalies also leads to this term. Witten showed that the 'hedgehog' solitons should be quantized as fermions for odd number of colour charges when the Wess–Zumino term is included [5]. In general, one uses a collective quantization approach to quantize the classical soliton solutions of the model [6]. The spin and isospin values of the quantum states pertaining to the 'hedgehog' soliton are given by $I = J = n/2$, $n = 1, 2, \ldots$ with $n = 1$ corresponding to the nucleons and $n = 3$ to the Δ resonance.

The static properties of the nucleon calculated from the Skyrme model are in agreement with the observed experimental values to an accuracy of 25–30%, in general [28]. However there exist some problems at the phenomenological level, some of which are serious.

There are soliton solutions with $B = n > 2$, which are cylindrically symmetric configurations in which the meson fields are twisted n times in the azimuthal plane [7, 8]. For these configurations, U has the form

$$U = \cos\Theta(\rho, z) + i\tau\cdot\hat{n}\sin\Theta(\rho, z), \tag{13}$$

$$\hat{n} = (\sin\alpha(\rho, z)\cos n\phi, \sin\alpha(\rho, z)\sin n\phi, \cos\alpha(\rho, z)), \tag{14}$$

where ρ, z and ϕ are the cylindrical coordinates. Θ and α satisfy the boundary conditions,

$$\Theta(\rho, z) \to 0 \quad \text{as} \quad \rho^2 + z^2 \to \infty, \Theta(\rho, -z) = \Theta(\rho, z), \Theta(0, 0) = \pi \tag{15}$$

and

$$\alpha(\rho, -z) = \alpha(\rho, z) + \pi. \tag{16}$$

The baryon number, which is also the winding number is given by

$$B = \frac{n\,\Theta(0, 0)}{\pi} = n. \tag{17}$$

There are classically stable solutions for this form of U with energy distributions corresponding to diffuse tori. Upon quantization, it is found that the lowest energy state has the quantum numbers of the deuteron and can be identified with it [8]. The binding energy is very high. The size of the soliton (measured by the root mean square radius) is nearly half the size of the physical deuteron.

The Skyrme model has been generalized to $SU(3)$. A straightforward generalization of the 'hedgehog ansatz' for $SU(2)$ to $SU(3)$ consists in taking

$$U = \begin{pmatrix} \cos\Theta + i\tau.\hat{r}\sin\Theta & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 1 \end{pmatrix}. \tag{18}$$

for the baryon octet [6]. Collective quantization will yield the baryon octet and the spin-$\frac{3}{2}$ decuplet, with the Wess–Zumino term playing a crucial role in giving the correct quantum numbers. The D/F ratio of the meson-baryon couplings is predicted correctly in the $SU(3)$ model [29]. However the mass spectrum of the baryons is not satisfactorily explained with the direct generalization of the 'hedgehog' [30]. There is an alternate approach to the strange baryons in the Skyrme model, in which a strange baryon is viewed as a bound state of a K-meson with the 'hedgehog' soliton [31]. This has some phenomenological advantages.

There is a hierarchy of 'spherically symmetric' ansatze, when we consider the Skyrme model for N_f number of flavours with $U \in SU(N_f)$ and $N_f \geq 3$ [6]. For general N_f, there are soliton solutions U characterized by the following generalized 'spherical' symmetry.

$$-i(\mathbf{r} \times \nabla)_i U + \left[\frac{\Lambda_i}{2}, U\right] = 0, \tag{19}$$

where Λ_i generate any $SU(2)$ or $SO(3)$ subgroup of the flavor $SU(N_f)$:

$$[\Lambda_i, \Lambda_j] = i\epsilon_{ijk}\Lambda_k. \tag{20}$$

In other words, Λ_i are the generators of the spin-J representation of $SU(2)$ or $SO(3) \subset SU(N_f)$. For a given value of N_f, J can take values $\frac{1}{2}, 1, ..., \frac{1}{2}(N_f - 1)$. For $N_f = 3$, apart from the solution with $\Lambda_i = \lambda_i/2$ $(i = 1, 2, 3)$ corresponding to the embedding of the $SU(2)$ hedgehog soliton in $SU(3)$ as in (20), there is another soliton solution corresponding to the following $SO(3)$ subgroup of $SU(3)$[9]:

$$\Lambda_1 = \lambda_7, \Lambda_2 = -\lambda_5 \quad \text{and} \quad \Lambda_3 = \lambda_2. \tag{21}$$

The baryon number for U corresponding to this is given by $B = 0, 2, 4 \cdots$. The lightest dibaryon in this sector is an $SU(3)$ singlet with strangeness $S = -2$ and with a mass of the order of 2.2 GeV, while the next set of states are octets with spin 1 or 2. These states have been predicted by the MIT bag model also [32].

When $N_f = 4$, apart from the two solitons characterized by $\Lambda_i = \lambda_i/2$ and the one described in equation (23), there is another soliton solution [10] corresponding to the following $SU(2)$ subgroup of $SU(4)$:

$$\Lambda_1 = \tfrac{1}{2}\sqrt{3}\lambda_1 + \lambda_6 + \tfrac{1}{2}\sqrt{3}\lambda_{13}, \quad \Lambda_2 = \tfrac{1}{2}\sqrt{3}\lambda_2 + \lambda_7 + \tfrac{1}{2}\sqrt{3}\lambda_{14},$$
$$\text{and} \quad \Lambda_3 = \tfrac{1}{2}\sqrt{3}\lambda_3 + \lambda_8 + \tfrac{1}{2}\sqrt{3}\lambda_{15}. \tag{22}$$

The associated soliton states have $B = 0, 1, 2, ...$ Upon collective quantization, it is found that the low-lying $B = 1$ states correspond to a 4^* representation of $SU(4)$ with spin-$\frac{3}{2}$, a 20^* with spins $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$ or $\frac{7}{2}$ and a $20^{*'}$ with spins $\frac{3}{2}$, $\frac{5}{2}$ or $\frac{9}{2}$ [10].

3. Baryonic solitons in the chiral quark-meson model

As has been mentioned already, the Skyrmion picture of baryons is troublesome from the point of view of QCD and the quark model. The favoured structure for the baryon involves three valance quarks bound by effective gluon exchanges, and interacting chiral-invariantly with meson fields. A model in which the quarks interact strongly with mesons only in a solitonic configuration may be considered as a simplified picture of the above.

In its simplest version, the chiral quark-meson model is described by the Lagrangian [33].

$$L_{\rm CQM} = i\overline{\Psi}\gamma^{\mu}\partial_{\mu}\Psi - gF_{\pi}(\overline{\Psi}_L U\Psi_R + \overline{\Psi}_R U^{\dagger}\Psi_L) + \frac{F_{\pi}^2}{4}{\rm Tr}(\partial_{\mu}U\partial^{\mu}U^{\dagger}) + \frac{F_{\pi}^2 m_{\pi}^2}{4}{\rm Tr}[(U+U^{\dagger})-2]. \tag{23}$$

Here $\Psi = \begin{pmatrix} u \\ d \end{pmatrix}$ denotes the isodoublet of quark fields u, d and $\Psi_{\rm L,R}$ are the left-handed and right-handed components,

$$\Psi_{\rm L,R} = \frac{1 \mp \gamma_5}{2}\Psi. \tag{24}$$

g is the quark-meson coupling constant, and as usual $U = \exp(-i(\boldsymbol{\tau}.\boldsymbol{\pi}/F_{\pi}))$. Except for the last term, the Lagrangian is invariant under the chiral $SU(2) \times SU(2)$ transformations given by

$$\Psi_{\rm L} \rightarrow A\Psi_{\rm L}, \quad \Psi_{\rm R} \rightarrow B\Psi_{\rm R}, \quad \text{and} \quad U \rightarrow AUB^{\dagger}, \tag{25}$$

where A and B are the left-handed and right-handed $SU(2)$ transformations. The model has the same structure as the nonlinear version of the σ-model introduced long ago by Gell–Mann and Levy [34], except that Ψ refers to the quark fields here, whereas in the old model it refers to the nucleons.

In the mean-field approximation, the quantum fluctuations of the π fields are ignored. Taking the cue from the Skyrme model, we assume a hedgehog ansatz for π corresponding to the $B = 1$ soliton solution with 3 quarks, i.e.,

$$U = \exp(-i\boldsymbol{\tau}.\hat{r}\,\Theta(r)) = \cos\Theta(r) - i\boldsymbol{\tau}.\hat{r}\sin\Theta(r). \tag{26}$$

(Here a negative sign is taken in the exponential to facilitate easy comparison with the literature on CQM). Anticipating the result, the following form for Ψ is chosen and it provides a self-consistent solution to the equations of motion:

$$\Psi = \begin{pmatrix} G(r)\chi \\ i\boldsymbol{\sigma}.\hat{r}F(r)\chi \end{pmatrix}\exp(-i\epsilon t), \tag{27}$$

where ϵ is the quark eigen energy and χ is a spin-isospin singlet, i.e.,

$$\chi = \frac{1}{\sqrt{2}}[|u\downarrow\rangle - |d\uparrow\rangle]. \tag{28}$$

This satisfies

$$(\boldsymbol{\tau} + \boldsymbol{\sigma})\chi = 0, \tag{29}$$

where $\boldsymbol{\tau}/2$ and $\boldsymbol{\sigma}/2$ are the isospin and spin generators.

With the ansatz given in (26)–(28), the equations of motion for G, F and Θ following from the space integral of L in (23) are [33]

$$G' + \epsilon F + gF_{\pi}(G\sin\Theta + F\cos\Theta) = 0, \tag{30}$$

$$F' + \frac{2}{r}F - \epsilon G + gF_{\pi}(G\cos\Theta - F\sin\Theta) = 0, \tag{31}$$

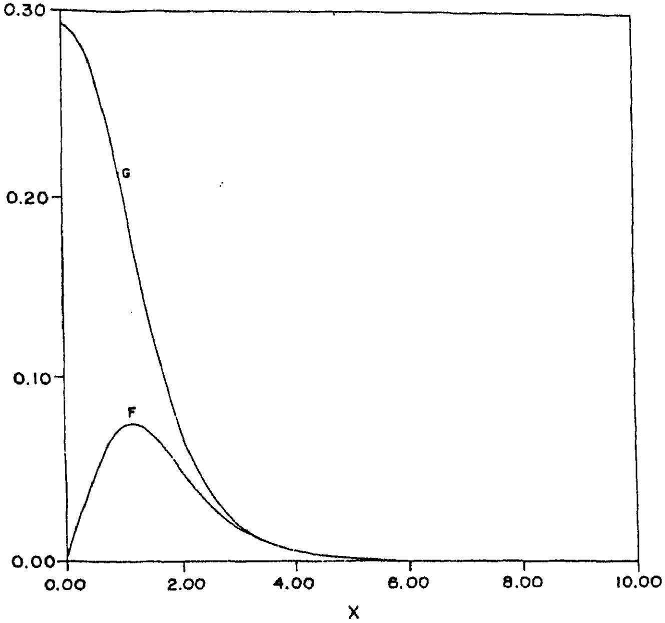


Figure 2. The upper and lower components $(G(x), F(x))$ of the quark function for the $B = 1$ soliton solution in the chiral quark-meson model.

$$\left(\Theta'' + \frac{2\Theta'}{r}\right) - \frac{\sin 2\Theta}{r^2} - m_\pi^2 \sin\Theta + \frac{3g}{F_\pi}[(G^2 - F^2)\sin\Theta + 2GF\cos\Theta] = 0. \tag{32}$$

The factor 3 arises in the last equation due to the fact that three quarks make up the soliton state.

We have solved equations (30)–(32) using the numerical code COLSYS subject to the boundary conditions $G(\infty) = F(\infty) = \Theta(\infty) = 0$, $F(0) = 0$ and $\Theta(0) = \pi$. [Note that $\sin\Theta(0) = 0$, to ensure that Θ is well-defined at the origin]. For g less than some critical value g_{crit}, no soliton solutions exist. This is not a surprising result and merely indicates that a certain minimum level of attraction is needed to bind the quarks. Numerically, it is found that $g_{\text{crit}} \approx 3.95$. One should choose g so that it leads to a reasonable energy for the soliton. It is found that $gF_\pi = 500\,\text{MeV}$ or $g = 5.367$ gives a satisfactory result.

It is convenient to use the dimensionless variables,

$$x = gF_\pi r, \quad G(x) = (gF_\pi)^{-3/2}G(r) \quad \text{and} \quad F(x) = (gF_\pi)^{-3/2}F(r). \tag{33}$$

In terms of these scaled variables, the normalization condition for the quark wavefunction is given by

$$4\pi\int_0^\infty x^2\,\mathrm{d}x(G^2(x) + F^2(x)) = 1. \tag{34}$$

The quark wavefunctions $G(x)$ and $F(x)$ are plotted as functions of x in figure 2. $\Theta(x)$ has essentially the same form as in figure 1. The size of the soliton ≈ 1.75, in dimensionless units. This corresponds to $\approx 0.7\,\mathrm{fm}$ as it should be.

The state corresponding to the above soliton is neither that of N nor that of Δ, but a superposition of both. One has to use appropriate projection techniques to find the quantum states corresponding to the neutron, proton and Δ states.

Till now, we have worked with the nonlinear σ model. One could also work with the linear σ-model [14] in which U has to be replaced by the matrix field

$$M = \frac{1}{F_\pi}(\sigma - i\tau\dot{\pi}). \tag{35}$$

The linear model is essentially given by (23), with U replaced by M and with the inclusion of some additional terms. However there is very little difference between the soliton solutions of the nonlinear and linear versions of the model. Phenomenologically, the linear model is on a slightly better footing (as it has more parameters) and leads to satisfactory results for the static properties of N and Δ.

There are classically stable cylindrically symmetric configurations with $B = 2$, that involve 6 quarks, in this model also, as in the Skyrme model. For these configurations U again has the form given in (13) and (14) (with $n = 2$), with the boundary conditions given by (15)–(16). For the $B = 2$ toroidal Skyrmion (corresponding to the deuteron),

$$(L_3 + 2T_3)|\,\text{physical state}\rangle = 0, \tag{36}$$

where L_3 and the T_3 are the z-components of the orbital angular momentum and the isospin generator respectively. This is simply the statement of cylindrical symmetry in the operational form, namely that a spatial rotation of α about the z-axis can be compensated for by an isorotation of -2α about the z-axis. For the quark fields in our model, the relevant symmetry operator is $K_3 = L_3 + T_3 + 2T_3$, where S_3 is the z-component of the spin angular momentum. As L_3 is integral and S_3 and T_3 are $\pm\frac{1}{2}$, it is obvious that the lowest values of K_3 are $\pm\frac{1}{2}$ for the quark fields. The most general form for Ψ with $K_3 = +\frac{1}{2}$ will be of the form:

$$\Psi = \begin{pmatrix} f_1 e^{-i\phi} u\uparrow + f_2 u\downarrow + f_3 e^{i\phi} d\uparrow + f_4 e^{2i\phi} d\downarrow \\ i(g_1 e^{-i\phi} u\uparrow + g_2 u\downarrow + g_3 e^{i\phi} d\uparrow + g_4 e^{2i\phi} d\downarrow) \end{pmatrix} e^{-i\epsilon t}, \tag{37}$$

where $f_1, f_2, f_3, f_4, g_1, g_2, g_3$ and g_4 are functions of ρ and z. Similarly, we can construct the most general Ψ with $K_3 = -\frac{1}{2}$. Plugging in this form for Ψ and U into the equations of motion, we get ten coupled nonlinear partial equations (f_i, g_i, Θ and α). It is not obvious that there exist any simple procedure to solve this set of equations. However, we assume a simplified ansatz for Ψ which satisfies all the symmetry requirements and yields the real variational minimum. We find that a spinor Ψ of the following form minimizes the energy

$$\Psi = \begin{pmatrix} G(\rho, z)\chi(\theta, \phi) \\ i\sigma\cdot\hat{r}F(\rho, z)\chi(\theta, \phi) \end{pmatrix} \exp(-i\epsilon t), \tag{38}$$

where χ is column vector in spin-isospin space which satisfies

$$(\tau\cdot\hat{n} + \sigma\cdot\hat{r})\chi = 0. \tag{39}$$

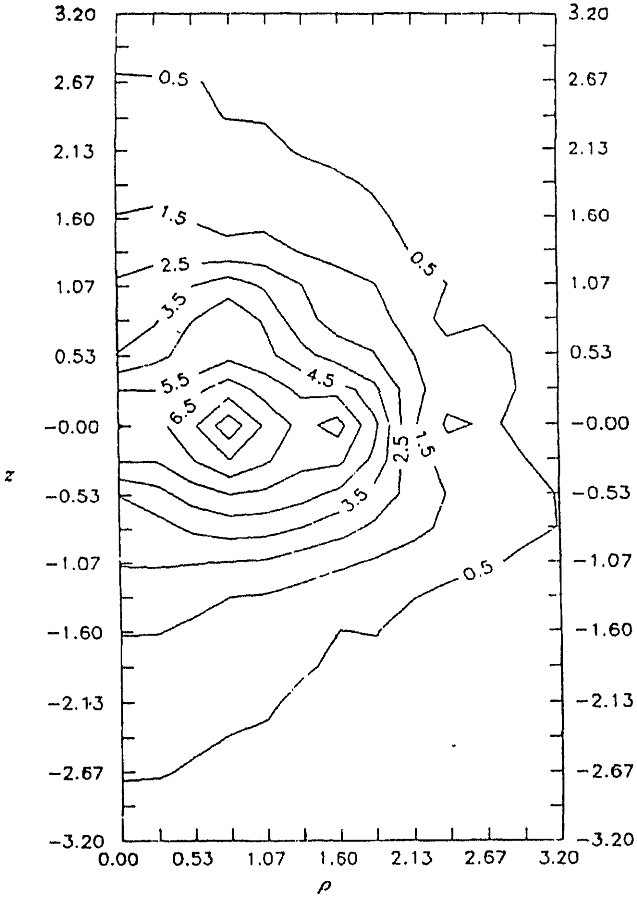


Figure 3. Contour plot of the energy density function $\varepsilon(\rho, z)$ for the $B = 2$ toroidal soliton in CQM.

The following two solutions χ_1 and χ_2 with $K_3 = +\frac{1}{2}$ and $-\frac{1}{2}$ respectively satisfy the above equation:

$$\chi_1 = \frac{1}{\sqrt{2(1 - \cos\theta\cos\alpha)}}\,[\sin\alpha u\downarrow - \sin\theta \mathrm{e}^{i\phi} d\uparrow - (\cos\alpha - \cos\theta)\mathrm{e}^{2i\phi} d\downarrow], \tag{40}$$

and

$$\chi_2 = \frac{1}{\sqrt{2(1 - \cos\theta\cos\alpha)}}\,[\sin\alpha d\uparrow - \sin\theta \mathrm{e}^{-i\phi} u\downarrow + (\cos\alpha - \cos\theta)\mathrm{e}^{-2i\phi} u\uparrow]. \tag{41}$$

It can be easily checked that the two solutions are degenerate in energy.

For the ansatz given by the equations (13), (14), (38) and (40), we have found that there is a solution by minimising the energy, using a relaxation procedure [15]. We find the

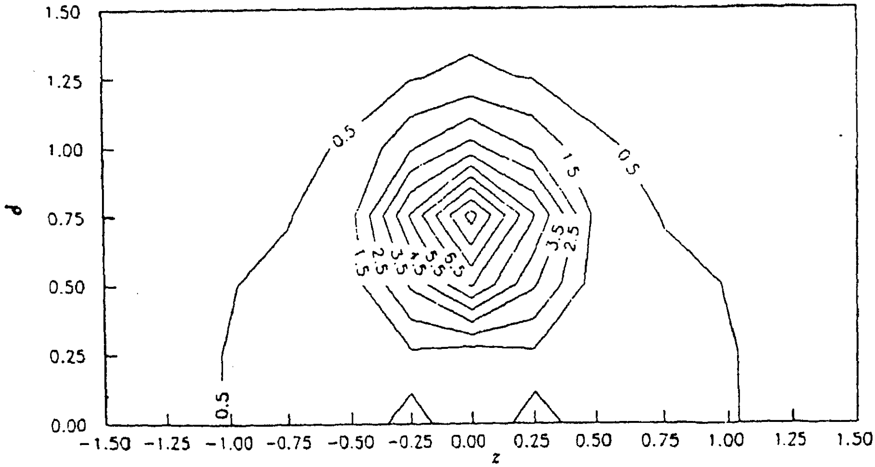


Figure 4. Contour plot of the energy density function $\varepsilon(\rho, z)$ for $B = 3$.

energy of the configuration to be 2019 MeV. As the classical energy for the $B = 1$ state is 1119 MeV, the classical binding energy is 219 MeV. In figure 3, we give the contour plot of the energy density function $\varepsilon(\rho, z)$ which clearly indicates a toroidal structure of the soliton solution, (the variables in the figure are scaled in units of gF_π). A naive value for the mean square radius of the charge distribution corresponding to the soliton would be given by $\langle r^2 \rangle = 2\pi \int \rho \mathrm{d}\rho \mathrm{d}z\, r^2\, (G^2 + F^2)$. We find $\langle r^2 \rangle^{1/2}$ to be 0.934 fm.

The symmetries of this toroidal solution closely parallel those of the corresponding object in the Skyrme model and its ground state has the quantum numbers of the deuteron, upon quantization. Whereas the stability of the solitons was due to the presence of the quartic term in the Skyrme model, here it is due to the quark-meson couplings.

For the solutions with $B = 3$ (9 quarks) and $B = 4$ (12 quarks) we have used a simplified ansatz of the form given in (38), where G and F are now taken to be functions of r only. Even this brings out the salient features of the soliton solution. By minimising the energy, we find that the soliton solutions are again toroidal in nature and closely parallel the corresponding toroidal solutions in the Skyrme model. We give the contour plot of the energy density $\varepsilon(\rho, z)$ for $B = 3$ in figure 4. The quark-meson model gives larger values for the binding energies. It is gratifying to note that the toroidal states are independent of the particular chiral-invariant model.

In the Skyrme model, the strange baryons are treated as excitations of a bound state of a kaon and the hedgehog soliton in the Callan–Klebanov approach [31]. In CQM, an equivalent of this method using a random phase approximation has been used with very limited success [35]. It would be natural to think of a strange baryon as corresponding to a soliton solution with nontrivial strange quark component and kaonic fields, with the s quark providing the strangeness quantum number just as the isospin of the baryon comes from the u/d quarks in this model. We do indeed find soliton solution with non-zero strangeness and $B = 1$ [36]. Unfortunately, these exist only for those values of the coupling constant g which are far greater than the phenomenologically acceptable number and hence do not correspond to the known strange baryons.

4. Regularity and chaos in the Skyrme model

In the following, we study the dynamical stability of the skyrmion. As discussed earlier our approach is different from the linear stability analysis: we explore the phase space around the soliton solution directly, by solving the time-dependent equations numerically [26]. The full nonlinearity of the theory comes into play in the time-evolution.

The spherically symmetric hedgehog ansatz in equation (10) is now made time-dependent:

$$U(\mathbf{r},t) = \cos\Theta(r,t) + i\tau\cdot\hat{r}\sin\Theta(r,t). \tag{42}$$

It can be seen that if U is of the hedgehog form initially, it will remain so at later times also, as the "spherical symmetry" is maintained. In other words, we are considering the restricted class of "spherically symmetric" fluctuations. Corresponding to this ansatz, the Lagrangian (integrated over space) is given by

$$\begin{aligned} L' = 4\pi\int r^2\,\mathrm{d}r\Bigg\{&\frac{F_\pi^2}{2}\left(\dot{\Theta}^2 - \Theta'^2 - \frac{2\sin^2\Theta}{r^2}\right)\\ &+\frac{1}{2e^2}\frac{\sin^2\Theta}{r^2}\left(2\dot{\Theta}^2 - 2\Theta'^2 - \frac{\sin^2\Theta}{r^2}\right) + \frac{m_\pi^2F_\pi^2}{4}(\cos\Theta - 1)\Bigg\}, \end{aligned} \tag{43}$$

where overdot denotes differentiation with respect to t. The Hamiltonian is given by

$$\begin{aligned} H' = 4\pi\int r^2\,\mathrm{d}r\Bigg\{&\frac{F_\pi^2}{2}\left(\dot{\Theta}^2 + \Theta'^2 + \frac{2\sin^2\Theta}{r^2}\right)\\ &+\frac{1}{2e^2}\frac{\sin^2\Theta}{r^2}\left(2\dot{\Theta}^2 + 2\Theta'^2 + \frac{\sin^2\Theta}{r^2}\right) + \frac{m_\pi^2F_\pi^2}{4}(1 - \cos\Theta)\Bigg\}. \end{aligned} \tag{44}$$

Θ now satisfies the following equation of motion:

$$\begin{aligned} &\left[\frac{x^2}{4} + 2\sin^2\Theta\right](\ddot{\Theta} - \Theta'') - \frac{x\Theta'}{2} + \sin 2\Theta\,(\dot{\Theta}^2 - \Theta'^2) + \frac{\sin 2\Theta}{4}\\ &\quad + \frac{\sin^2\Theta\sin 2\Theta}{x^2} + \frac{x^2\beta}{4}\sin\Theta = 0, \end{aligned} \tag{45}$$

where x and β have been defined earlier.

We consider the system in a finite volume. This is perfectly justified, as the skyrmion has a finite size. We discretize the spatial variable x and replace the derivatives by corresponding difference expressions:

$$\Theta' = \frac{\Theta(i+1,t) - \Theta(i,t)}{\delta x} \tag{46}$$

and

$$\Theta'' = \frac{\Theta(i+1,t) - 2\Theta(i,t) + \Theta(i-1,t)}{(\delta x)^2}, \tag{47}$$

the number of lattice points. In our analysis, we have taken the length of the system to be 8 units(the size of the skyrmion in dimensionless unit $x \approx 2$) and the number of lattice points to be 64.

We write $\Theta(x,t)$ in the following form

$$\Theta(x,t) = \Theta_0(x) + \delta\Theta(x,t), \tag{48}$$

where Θ_0 represents the static solution and the fluctuation $\delta\Theta(x,t)$ is a combination of various harmonic modes of the form

$$\delta\Theta(x,t) = \sum_{j=1}^{N-1} a_j(t) \sin\left(\frac{j\pi x}{L}\right). \tag{49}$$

Here, L is the length of the system (8 in our case) and N is the number of lattice points (64 for us). $a_j(t)$ is the amplitude of the jth mode which is got by inverting the series given in (49):

$$a_j(t) = \frac{2}{L}\int_0^L \delta\Theta(x,t) \sin\left(\frac{j\pi x}{L}\right) \mathrm{d}x. \tag{50}$$

This way, the problem is converted into one corresponding to a system of coupled anharmonic oscillators and one can study the time evolution of the amplitudes of the modes. This is essentially the Fermi–Pasta–Ulam approach for coupled systems [21].

As U has to be well-defined at $r = 0$ and $U \to 1$ as $r \to \infty$, we have the boundary conditions

$$\delta\Theta(0,t) = \delta\Theta(L,t) = 0 \tag{51}$$

which is automatically satisfied by the harmonic expansion for $\delta\Theta$ in (49). Moreover, we take

$$\dot{\Theta}(x,0) = 0. \tag{52}$$

Also, we excite only the $N/2$th or 32nd mode initially, following standard practice.

Thus,

$$a_j(0) = 0 \qquad j \neq 32. \tag{53}$$

$A \equiv a_{32}(0)$ is our perturbation parameter.

We now integrate out the equation of motion in (45) with the aforementioned initial conditions using a standard Runge–Kutta routine with a time step 0.001 and 5×10^5 iterations, i.e., real time $t = 500$. The total energy given by H' in equation (44) was constant to an accuracy of 0.2% for this time step. We have studied the time evolution of the system for various values of the perturbation parameter A.

We have plotted the fluctuation $\delta\Theta(x,t)$ at $t = 0$ and 500 for $A = 0.1$ in figure 5. Apart from large fluctuations near $x = 0$, $|\delta\Theta(x,t)| \approx 0.1$ in this case. This is understandable as the skyrmion dynamics is dominated by the small x-region and the deviation from the

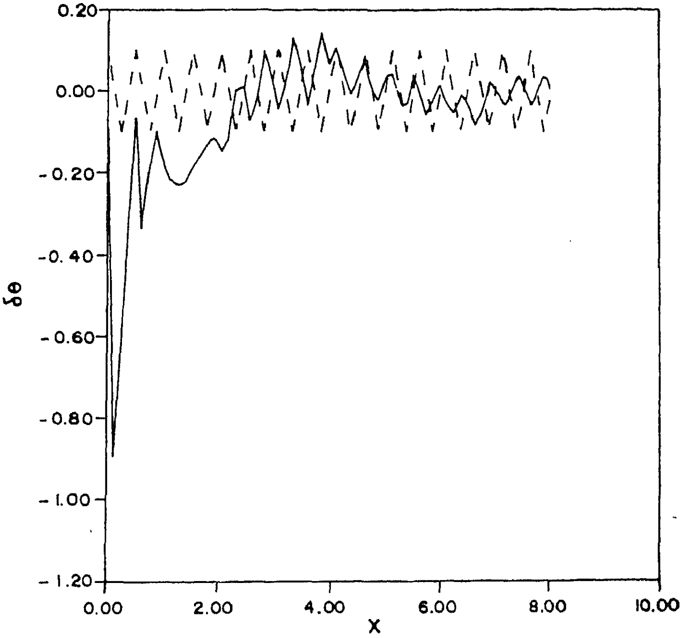


Figure 5. $\delta\Theta(x,t)$ at $t=0$ (dotted line) and $t=500$ (continuous line) for $A=0.1$, in the Skyrme model.

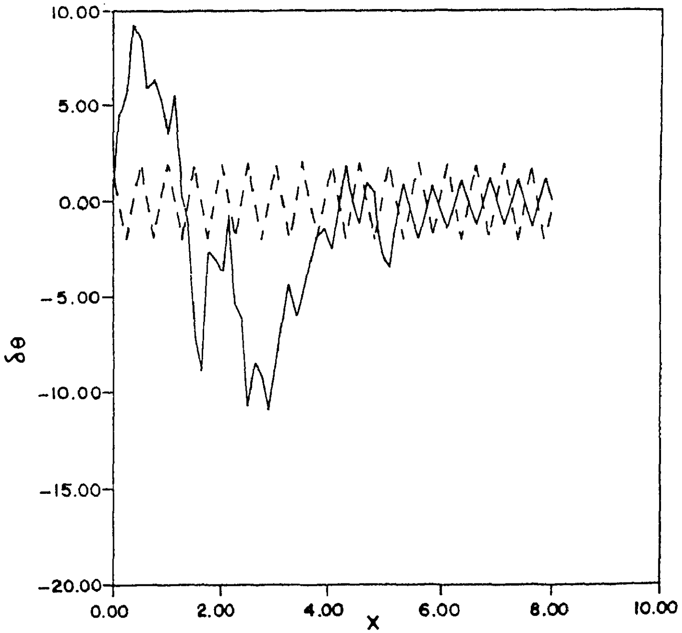


Figure 6. $\delta\Theta(x,t)$ at $t=0$ and $t=500$ for $A=2$.

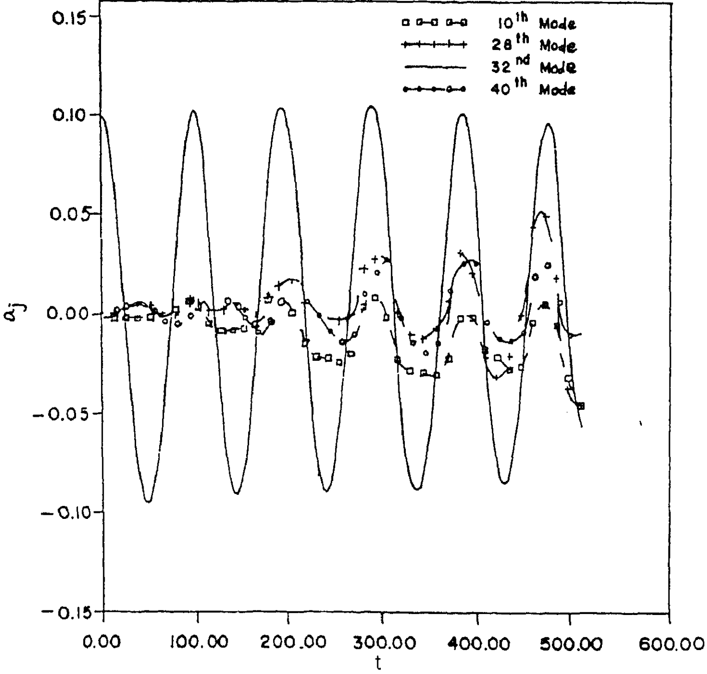


Figure 7. Amplitudes $a_j(t)$ of the modes corresponding to $j = 10, 28, 32$ and 40 plotted against time for $A = 0.1$ in the Skyrme model.

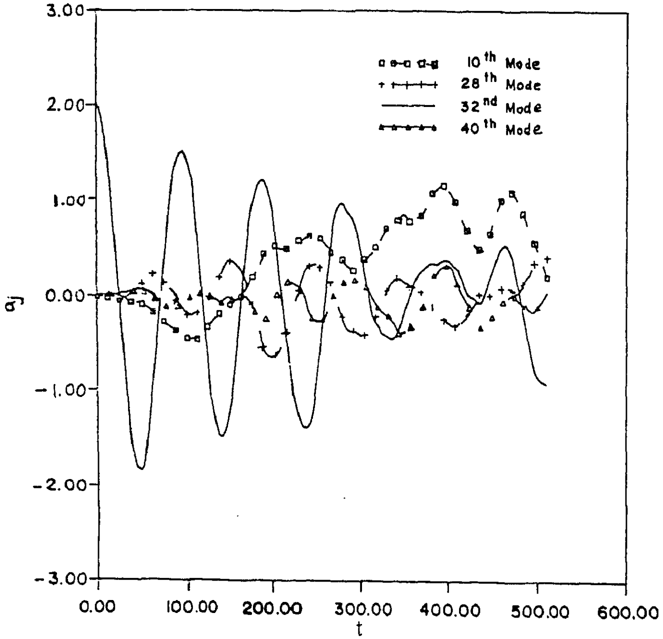


Figure 8. Amplitudes $a_j(t)$ for $A = 2$.

skyrmion is small when $A = 0.1$. Contrast this with the behaviour of $\delta\Theta(x,t)$ for $A = 2$ in figure 6. Now $\delta\Theta(x,t)$ has violent fluctuations for substantially longer intervals in x. For $A = 2$, the initial fluctuation is of the same order of magnitude as the static solution itself and after a long time, Θ would have no resemblance to the skyrmion!

The behaviour of the amplitudes of various harmonic modes is an important criterion for regularity or chaos. We have plotted the amplitudes $a_j(t)$ for $j = 10$, 28, 32 and 40 as functions of time for $A = 0.1$ and 2 in figures 7 and 8 respectively. For $A = 0.1$, the amplitude of the (initial) 32nd mode has an extremely regular periodic behaviour (note that it does not remain constant!). However the amplitudes of the other modes increase gradually on the average. The energy transfer to other modes is expected to be substantial after a longer interval of time. In contrast to this, for $A = 2$, the amplitude of the initial mode decreases gradually on the average, though there is a periodicity in the variation. After $t = 200$, other modes are excited substantially. It is to be expected that after a longer interval of time, all the modes would be of same magnitude leading to 'thermalization' and 'spatio-temporal chaos'.

Finally, we have plotted the root mean square radius of the soliton, normalized to its initial value, as a function of time for $A = 0.1$ and 2 in figures 9 and 10. For $A = 0.1$ the radius of the soliton has a periodic behaviour, whereas for $A = 2$, the radius increases gradually on the average.

We have increased the spatial resolution by a factor of 2 by taking the number of lattice point, N, to be 128 and checked that there is an insignificant change in the numerical results.

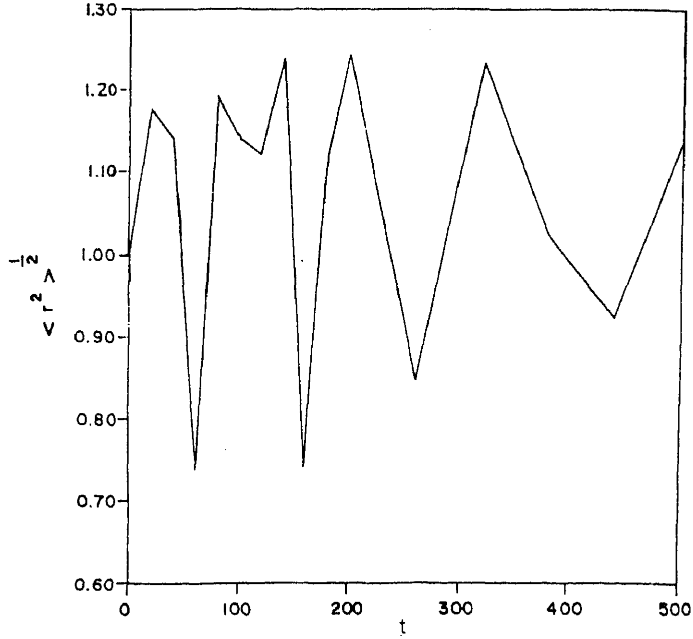


Figure 9. Variation of the root mean square radius $\langle r^2 \rangle^{1/2}$ (normalized to unity at $t = 0$), with time for $A = 0.1$ in the Skyrme model.

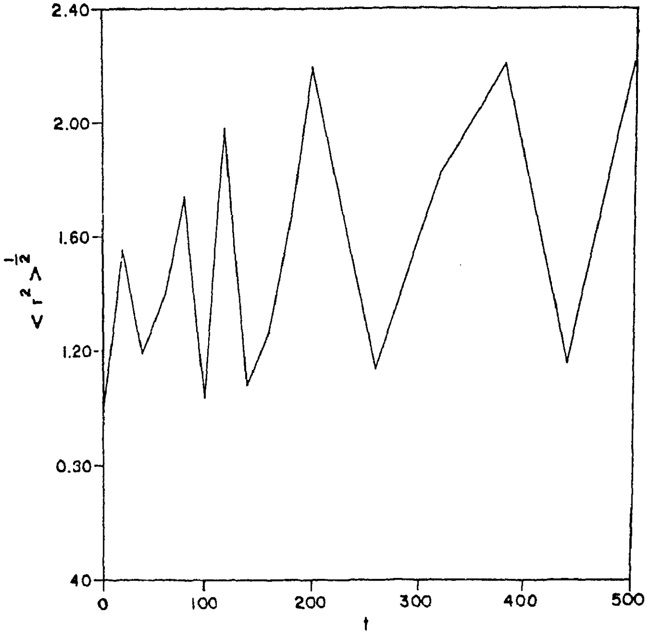


Figure 10. Variation of $\langle r^2 \rangle^{1/2}$ for $A = 2$.

Hence, when the strength of the initial fluctuations is small, the evolution of the system is somewhat regular atleast for $t \approx 500$; however, there are already hints that the fluctuations would become random after a longer interval of time. If the strength of the perturbation is increased, energy transfer to other modes (different from the initial one) is substantial and the fluctuations are now chaotic. This is a clear indication that the phase space near the Skyrme soliton exhibits spatio-temporal chaos and the system undergoes "thermalization" after a sufficiently long time. The system is non-integrable. There is no evidence of any phase transition in the system.

It might be thought that the above behaviour is characteristic of any non-linear system. As the Skyrme model is a highly non-linear system with an interaction term which has four space-time derivatives of the field, this could have been anticipated. However, it should be remembered that the skyrmion is a topological object. The topological conservation law imposes its own constraint on the dynamics. As the fluctuations satisfy the boundary condition $\delta\Theta(0,t) = \delta\Theta(\infty,t) = 0$ at all times, the baryon number is not affected by the dynamics. Our work has revealed that within the constraint of the topological conservation law, the soliton exhibits a chaotic behaviour after a sufficiently long time, for large enough perturbations.

5. Stability of the $B = 1$ soliton of the chiral quark-meson model

In this section, we investigate the dynamical behaviour of the chiral quark-meson model along the same lines as in § 4 [26].

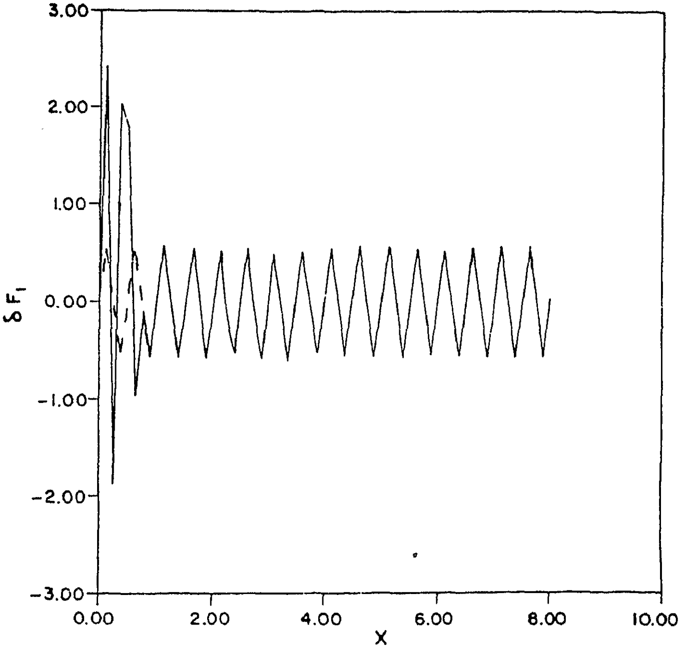


Figure 11. $\delta F_1(x,t)$ at $t = 0$ (dotted line) and $t = 500$ (continuous line) for $A = 2$ in the chiral quark-meson model. The two curves merge after some value of x.

In the time-dependent case, for the $B = 1$ soliton, we take:

$$U = \cos\Theta(r,t) - i\tau\cdot\hat{r}\sin\Theta(r,t) \tag{54}$$

$$\Psi = \begin{pmatrix} [G_1(r,t) + iG_2(r,t)]\chi \\ i\sigma\cdot\hat{r}[F_1(r,t) + iF_2(r,t)]\chi \end{pmatrix} \exp(-i\epsilon t) \tag{55}$$

for a consistent set of equations.

The equations of motion following from L_{CQM} in (23) would now be:

$$\dot{G}_1 = -F_2' - \frac{2F_2}{x} - (-\epsilon + \cos\Theta)G_2 - \sin\Theta F_2 \tag{56}$$

$$\dot{F}_1 = -G_2' - (\epsilon + \cos\Theta)F_2 - \sin\Theta G_2 \tag{57}$$

$$\dot{G}_2 = -F_1' - \frac{2F_2}{x} - (-\epsilon + \cos\Theta)G_1 + \sin\Theta F_1 \tag{58}$$

$$\dot{F}_2 = G_1' + (\epsilon + \cos\Theta)F_1 + \sin\Theta G_1 \tag{59}$$

$$\begin{aligned} &\ddot{\Theta} - \Theta'' - \frac{2\Theta'}{x} + \frac{\sin 2\Theta}{x^2} \\ &- 3g^2\{\sin\Theta(G_1^2 + G_2^2 - F_1^2 - F_2^2) \\ &+ 2\cos\Theta(G_1F_1 + G_2F_2)\} - \beta^2\sin\Theta = 0, \end{aligned} \tag{60}$$

where the scaled variables x, $G_i(x)$ and $F_i(x)$ are defined in (33) and prime denotes differentiation with respect to x. As in the case of the Skyrme model, we take the length

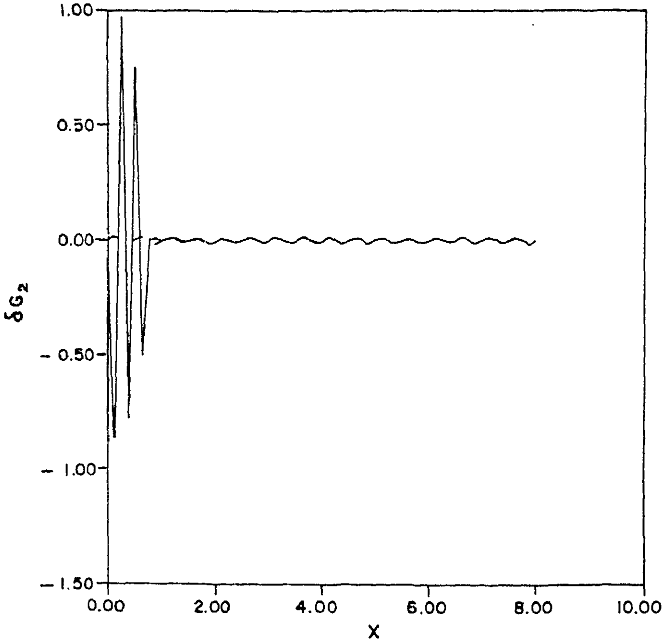


Figure 12. $\delta G_2(x,t)$ at $t=0$ and $t=500$ for $A=2$.

of the system to be 8 and the number of lattice points to be 64. The fields at any time can be written as

$$\Psi(x,t) = \Psi_{,0}(x,t) + \delta\Psi(x,t), \tag{61}$$

where

$$\delta\Psi(x,t) = \sum_{j=0}^{N-1} a_j^{\Psi}(t) \sin\left(\frac{j\pi x}{L}\right). \tag{62}$$

Here Ψ stands for the fields G_1, F_1, G_2, F_2 or Θ and $\Psi_{,0}$ is the static solution for the corresponding field. ($G_{2,0} = F_{2,0} = 0$). $a_j^{\Psi}(t)$ represents the amplitude of the jth mode of the fluctuation in field Ψ. At any time $a_j^{\Psi}(t)$ can be found by inverting the series in (62), i.e.,

$$a_j(t) = \frac{2}{L}\int_0^L \delta\Psi(x,t) \sin\left(\frac{j\pi x}{L}\right) \mathrm{d}x. \tag{63}$$

We take

$$\dot{\Psi}(x,0) = 0 \tag{64}$$

as in the previous section. We excite the 32nd mode initially for Θ, G_1 and F_1, G_2 and F_2:

$$a_{32}^{\Theta}(0) = A, \tag{65}$$

$$a_{32}^{G_1}(0) = AG_{1,0}(x=0), \tag{66}$$

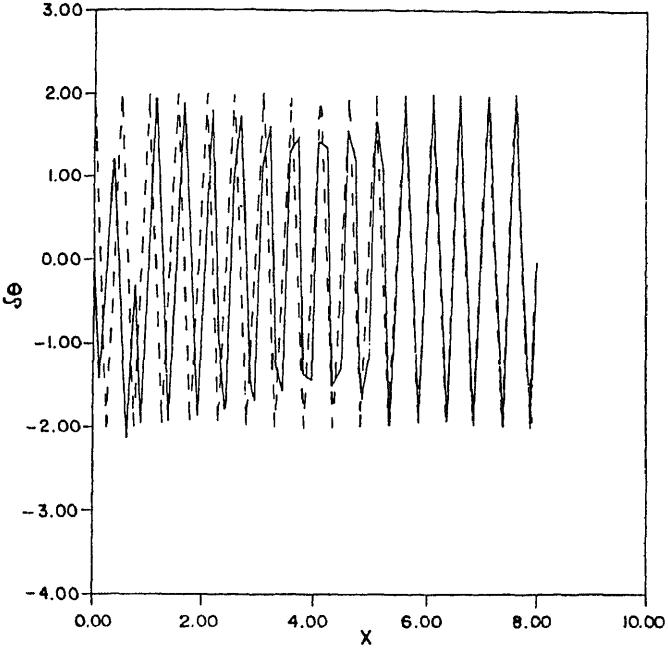


Figure 13. $\delta\Theta(x,t)$ at $t = 0$ (dotted line) and $t = 500$ (continuous line) for $A = 2$ in the chiral quark-meson model.

$$a_{32}^{F_1}(0) = AG_{1,0}(x = 0), \tag{67}$$

$$a_{32}^{G_2}(0) = a_{32}^{F_2} = AF_{1,0}(x = 1.1), \tag{68}$$

where A is the perturbation parameter. In eqs. (66)–(68), we have the factor $G_{1,0}(0)$ and $F_{1,0}(x = 1.1)$ to ensure that the perturbation has the natural scale for G_1 , F_1, G_2 and F_2.

We now integrate the 5 coupled equations (56) to (60) using the standard Runge–Kutta routine with a time step 0.001 and 5×10^5 iterations, i.e., real time $t = 500$. The time evolution of the system is numerically determined for various values of the amplitudes.

We have plotted the fluctuations δF_1, δG_2 and $\delta\Theta(x,t)$ for $t = 0$ and 500 for $A = 2$ in figures 11–13. We observe that there is very little variation in these functions except at the origin even though the amplitude is large. This is strikingly different from the time evolution of the fluctuation around the soliton in the Skyrme model, for the same value of the amplitude of the 32nd mode.

We have plotted the amplitudes of the modes $j = 10,28, 32$ and 40 corresponding to the functions G_1, F_1 and Θ in figures 14–16. Again there is very little variation in the amplitude of the initial 32nd mode even after $t = 200$. The amplitudes of all the other modes remain very small even after a long time, for all the functions.

Hence, for reasonably large fluctuations, the system exhibits a very regular behaviour. It is thus clear, that the soliton of the chiral quark-meson model is much more stable than the Skyrmion for comparable strengths of fluctuations.

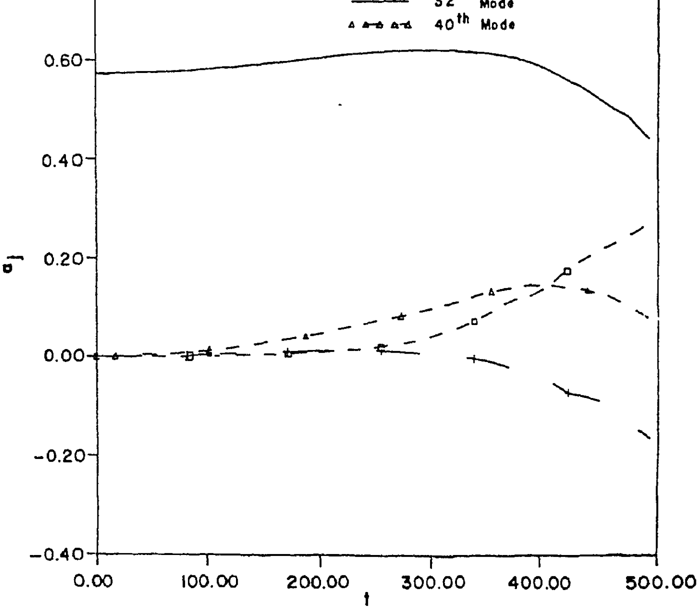


Figure 14. Amplitudes $a_j(t)$ of the function G_1 corresponding to $j = 10, 28, 32$ and 40 plotted against time for $A = 0.1$ in CQM.

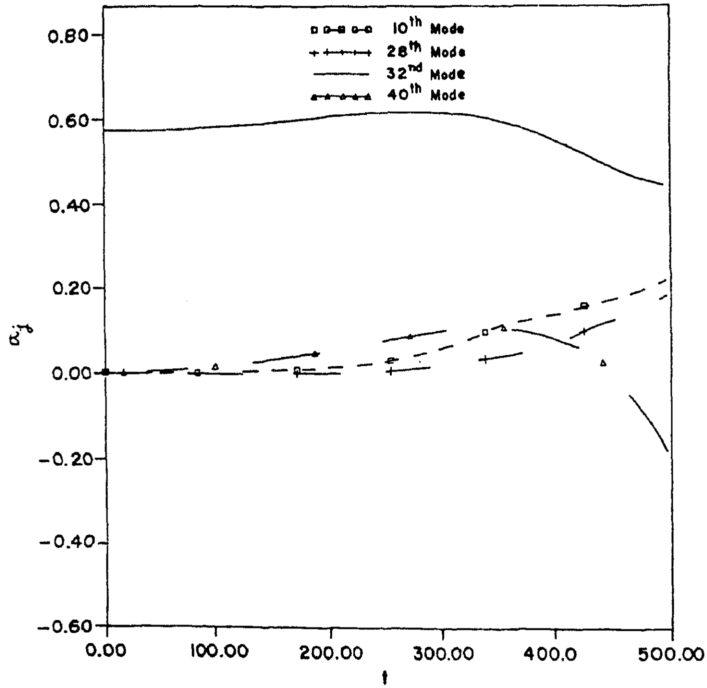


Figure 15. Amplitudes $a_j(t)$ of the function F_1 for $A = 2$.

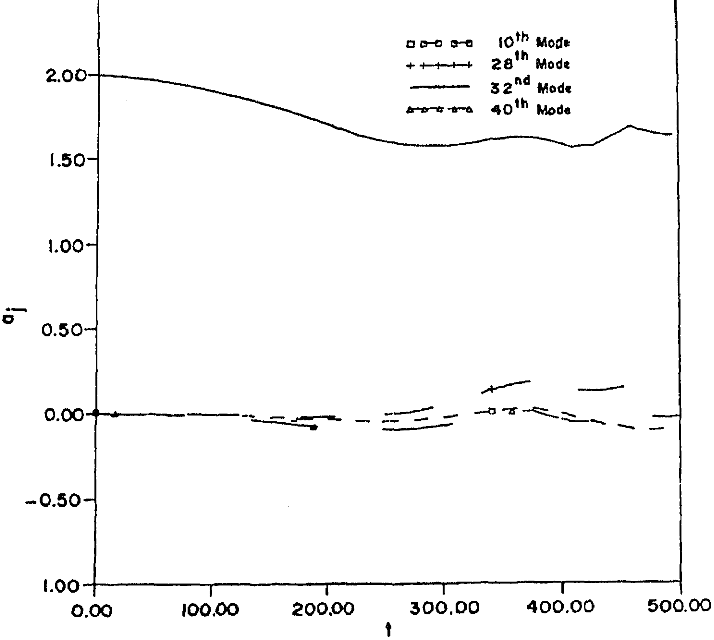


Figure 16. Amplitudes $a_j(t)$ of the function Θ for $A = 2$, in CQM.

6. Discussion

In this article, we have described the 'soliton solutions' of two nonlinear chiral models, namely, the Skyrme model and the chiral quark meson model and their identification with the baryons, the fundamental constitutents of matter. It is fascinating that a fully bosonic theory like the Skyrme model has topological soliton solutions which are identified with fermions. This is because of the presence of the topological Wess–Zumino term and the collective quantization procedure, discussion of which is beyond the scope of this review. The interactions between the solitons of the two models has been studied and it has been shown that the gross features of the internucleon potential are reproduced. The scattering properties of these solitons in $3 + 1$ dimensions is yet to be investigated. However, there are some very interesting results on soliton scattering in the Skyrme model in $2 + 1$ dimensions. Peyrard, Piette and Zakrzewski have shown that in scattering processes initiated at low relative velocities, the Skyrmions bounce back while at velocities larger than some critical value, they scatter at right angles [37]. Moreover, the Skyrmions preserve their shape after collision. This seems to be a characteristic feature of solitons of any S^2 model in $2 + 1$ dimension. There are indications that in all head-on collisions of N indistingushable soliton like objects in $2 + 1$ dimensions, the scattering angle is π/N [38]. It would be of great interest to study the scattering of Skyrmions and related topological objects in $3 + 1$ dimensions to find whether they retain their shape and size and whether the scattering angles are still discretized.

We have considered another aspect of the soliton solutions in these models, namely stability with respect to fluctuations. It is known that non-abelian gauge theories with and without the Higgs fields exhibit spatio-temporal chaos. We have estabilished that the Skyrme model, an effective Lagrangian theory based on QCD, also shares this feature, by direct numerical computations. It is interesting to note that topological stability and spatio-temporal chaos coexist in the model. In contrast to this, the chiral quark-meson model which is also QCD based is very stable for comparable strengths of fluctuations around the soliton solutions and comparable time scales. The region of phase space around the soliton solution of the model exhibits more regularity, in contrast to the phase space around the Skyrme soliton. The reason perhaps lies in the fact that the couplings in the Skyrme model are more nonlinear due to the four-derivative terms. In the chiral quark meson model, the four-derivative term is absent and instead of that we have the Yukawa coupling term. It would be interesting to study the dynamical behaviour of other field theories with fermion fields to find whether they are also very stable.

The large fluctuations around the chiral soliton solutions that we have discussed would be of importance when we consider interactions among them. Chaotic field configurations have been known to play a role in multiparticle hadron production processes; indeed a chaoticity parameter is introduced to explain the average multiplicity[39]. Our work is a preliminary exercise in demonstrating concretely how spatio-temporal chaos arises in field-theoretic models of strong interactions. It remains to be seen how specific physical processes are influenced by it.

References

[1] See for example, R Rajaraman, Solitons and Instantons (North Holland Publishing Company, Amsterdam, 1982) and references therein
J Hong, Y Kim and P Y Pac, *Phys. Rev. Lett.* **64**, 2230 (1990)
R Jackiw and E J Weinberg, *Phys. Rev. Lett.* **64**, 2234 (1990)

[2] W A Bardeen, S D Drell, M Weinstein and T M Yan, *Phys. Rev.* **D11**, 1094 (1975)

[3] R Friedberg and T D Lee, *Phys. Rev.* **D16**, 1096 (1977); **D17**, 2623 (1978)

[4] T H R Skyrme, *Proc. R. Soc. Am.* **A260**, 127 (1961)
V G Makhankov, Y P Ryabkov, and V I Sanyuk, *The Skyrme Model* (Springer Verlag, Berlin, Heidelberg, 1993)

[5] E Witten, *Nucl. Phys.* **B223**, 433 (1983); in this paper, essential use is made of the Wess–Zumino term given by J Wess and B Zumino, *Phys. Lett.* **B37**, 95 (1971)

[6] A P Balachandran, *Lectures at 1985 TASI* edited by M Bowick and F Gursey (World Scientific, Singapore, 1986)

[7] V B Kopeliovich and B E Shtern, *JETP. Lett.* **45**, 203 (1987)

[8] E Braaten and L Carson, *Phys. Rev.* **D38**, 3525 (1988)

[9] A P Balachandran, A Barducci, F Lizzi, V G J Rodgers and A Stern, *Phys. Rev. Lett.* **52**, 887 (1984)

[10] Faheem Hussain and M S Sriram, *Phys. Rev. Lett.* **55**, 1169 (1985)

[11] M P Mattis and M Karliner, *Phys. Rev.* **D31**, 2833 (1985)
M P Mattis and M Peskin, *Phys. Rev.* **D32**, 58 (1985)

[12] A Jackson, A D Jackson and V Pasquier, *Nucl. Phys.* **A432**, 562 (1985)

[13] R D Ball, *Int. J. Mod. Phys.* **A23**, 4391 (1990)

[14] S Kahana, G Ripka and V Soni, *Nucl. Phys.* **A415**, 351 (1984)
M C Birse and M K Banerjee, *Phys. Rev.* **D31**, 118 (1984)

[15] J Segar, M Sripriya and M S Sriram, *Int. J. Mod. Phys.* **E3**, 769 (1994)
J Segar, M Sripriya and M S Sriram, *Phys. Lett.* **B342**, 201 (1995)

[16] S G Matinyan, G K Savvidy and N G T Savvidy, *Zh. Eksp. Teor. Fiz.* **80**, 830 (1981) [*Sov. Phys. JETP.* **53**, 421 (1981)]

[17] A Giansanti and P D Simic, *Phys. Rev.* **D38**, 1352 (1988)
C N Kumar and A Khare, *J. Phys.* **A22**, L849 (1989)
B Dey, C N Kumar and A Sen, *Int. J. Mod. Phys.* **A8**, 1755 (1993)
B A Bambah, S Lakshmibala, C Mukku and M S Sriram, *Phys. Rev.* **D47**, 4677 (1993)
M S Sriram, C Mukku, S Lakshmibala and B A Bambah, *Phys. Rev.* **D49**, 4246 (1994)
S Lakshmibala, in *Computational aspects in chaos and nonlinear dynamics* edited by G Ambika and V M Nandakumaran (Wiley Eastern, New Delhi, 1994) pp. 37–47

[18] C Mukku, M S Sriram, J Segar, B A Bambah and S Lakshmibala, submitted for publication
S Lakshmibala, B A Bambah, M S Sriram and C Mukku, *Pramana – J. Phys.* **48**, 617 (1997)

[19] M S Sriram, in *Computational aspects in chaos and nonlinear dynamics* edited by G Ambika and V M Nandakumaran (Wiley Eastern, New Delhi, 1994) pp. 37–47

[20] S G Matinyan, E B Prokhorenko, and G K Savvidy, *Pis'ma Zh. Eksp. Teor. Fiz.* **44**, 109 (1986) [*JETP Lett.* **44**, 138 (1986)]

[21] E Fermi, J R Pasta and S M Ulam, *Los Alamos Sci. Lab. Tech. Report No. LA-1940* (1955) (unpublished)

[22] M P Joy and M Sabir, *J. Phys.* **A22**, 5153 (1989)
M P Joy and M Sabir, *Pramana – J. Phys.* **38**, L91 (1992)

[23] T Kawabe and S Ohta, *Phys. Rev.* **D41**, 1983 (1990)

[24] T Kawabe and S Ohta, *Phys. Rev.* **D44**, 1274 (1991)

[25] V G Makhankov *et al*, in ref [4]

[26] J Segar and M S Sriram, *Phys. Rev.* **D53**, 3976 (1996)

[27] G Adkins and C Nappi, *Nucl. Phys.* **B233**, 109 (1984)

[28] G S Adkins in *Chiral Solitons* edited by K F Liu (World Scientific, Singapore, 1987)

[29] M S Sriram, H S Mani and R Ramachandran, *Phys. Rev.* **D30**, 1141 (1984)

[30] E Guadagnini, *Nucl. Phys.* **B236**, 35 (1984) see also ref [29]

[31] C G Callan and I Klebanov, *Nucl. Phys.* **B262**, 365 (1985)

[32] R L Jaffe, *Phys. Rev. Lett.* **38**, 195 (1977)

[33] J Segar, Baryonic solitons in quark-meson models of strong interaction, Ph.D. Thesis, University of Madras (1994)

[34] M Gellmann and M Levy, *Nuovo Cimento* **16**, 705 (1960)

[35] J A McGovern and M C Birse, *Nucl. Phys.* **A506**, 392 (1993)

[36] J Segar and M S Sriram, *Pramana – J. Phys.* **40**, 291 (1993)

[37] M Peyrard, B Piette and W Zakrzewski, *Nonlinearity* **5**, 565, 585 (1992)

[38] A Kudryavtsev, B Piette and W J Zakrzewski, *Phys. Lett.* **A180**, 119 (1993)

[39] P Carruthers, E M Friedlander and R M Weiner, *Physica* **D23**, 138 (1986)
P Carruthers, E M Friedlander, C C Shih and R M Weiner, *Phys. Lett.* **222**, 487 (1989)

Control methods for problems of mixing and coherence in chaotic maps and flows

ASHUTOSH SHARMA and NEELIMA GUPTE*
Department of Physics, University of Pune, Pune 411 007, India
*Department of Physics, Indian Institute of Technology, Madras 600 036, India

Abstract. We review a variety of control methods which are capable of enhancing the chaoticity and mixing properties of chaotic flows and also methods which work towards promoting the coherence properties of such flows. We discuss a parameter control method which can enhance the chaoticity and the rate of mixing of dissipative as well as conservative flows and outline methods which promote global mixing by the addition of noise and by preventing island formation. As the inverse side of this problem, we summarize methods which can create coherent structures in chaotic dynamical flows. We also discuss the utility of these methods from the point of view of applications as well as for understanding phenomena which occur in natural systems.

Keywords. Lagrangian chaos; extended systems; control.

PACS No. 5.45

1. Introduction

The control of chaos has been a topic of much recent interest [1]. The control of chaos has found interesting applications in contexts ranging from lasers, mechanical systems, and electronic circuits, to chemical reactions and biological systems. Some of this effort has been directed towards the imposing of periodic orbits on chaotic systems, other directions of effort have been towards maintaining chaos in contexts where the destruction of chaotic behaviour is undesirable. The enhancement of chaos via control methods can also find useful applications. We discuss control methods which work towards enhancing chaos in chaotic flows in this review. We also discuss methods which work towards creating coherence in chaotic flows.

The enhancement of chaos can be desirable and useful in a variety of situations. In the case of biological systems, there are several instances of situations where maintaining or enhancing chaos is desirable. It has been suggested that the pathological destruction of chaotic behaviour may be responsible for heart failure [2], and some types of brain seizures [3]. Techniques which are capable of enhancing and maintaining chaos could be useful in such contexts [4].

An important example of a situation where enhancing chaos is useful, is the process of mixing [5, 6]. The mixing properties of a flow are important in various engineering and industrial problems like combustion processes [7], viscous liquids [8], chemical reactions [9, 10], heat transfer processes [11]. Often the performance of a particular process, its yield, the rates and even the product depends on the mixing properties. For example, a well mixed fuel-air mixture can give rise to efficient combustion and a chemical reaction in which the reactants are better mixed enhances the rate of reaction and the yield. Therefore it is essential to look at the underlying dynamical properties which give rise to efficient, or poor mixing. Mixing is a consequence of the stretching and folding of chaotic flows. A system which has exponential stretching, as in a chaotic flow, can mix efficiently. Many mixing processes like fluid flows, combustion processes, chemical reactions, heat transfer processes, etc can be modelled by chaotic flows [6]. An enhancement of the chaoticity of such systems can lead to an enhancement of the rate of mixing, an outcome which has desirable consequences in many of the situations described above.

Coherent structures, on the other hand are the flip side of the coin. These are regions in the flow which remain unmixed and may or may not permit the exchange of trajectories with globally mixing surrounding. The presence of a coherent structure in otherwise strongly mixing surroundings is a puzzling but ubiquitous feature of natural systems. Examples of coherent structures can be found in a variety of natural contexts, ranging from planetary [12] and geophysical [13] systems, to chemical mixers [14], coupled oscillator arrays [15] and turbulent flows [16]. Again, low dimensional chaotic flows have proved to be capable of encapsulating much of the generic behaviour seen in such systems [17–19].

It is therefore interesting to inquire into the reasons behind the formation of coherent regions created in such chaotic flows. Control methods capable of the creation of coherent structures in chaotic flows can provide insights into these reasons. These structures could also find useful applications. A very interesting application of such structures has been proposed recently. Such structures could be utilised to control the rate, products and yield of chemical reactions by using them as controlled release capsules [20]. The idea is to release one or more of the reactants in a chemical reaction at a specific time or to control the concentration of such reactants. These reactants are placed on coherent structures and the flow is mixed. The coherent structure prevents exchange of these reactants with the surrounding environment. Then at a desired time the structure can be destroyed by application of noise and the reactants released into the reaction. Such a capsule can be used in a continuous flow reactor and permits the control of various reaction rates and products.

The existence of mixing and coherent regions in flows can have relevance for a variety of other phenomena. The flow region, as visualized by injecting a passive scalar such as a dye, acts as a basic template for various interesting physical processes such as aggregation, coagulation or breakup [21, 22]. The fabric of the underlying chaotic dynamics provides a template on which various processes like aggregation, coagulation and drop breakup occurs. By exploiting the properties and symmetries of this fabric, the coherence and mixing properties of mixing flows, one can understand and manipulate these processes better. Here again, one can show that the shape and sizes of various aggregates and rates of various processes depend on mixing properties and presence of coherent structures.

It is clear that many of the situations which constitute our motivation correspond to spatially extended systems such as fluid flows. It is therefore relevant to discuss the sense in which the analysis of the dynamics of low-dimensional systems in phase space is are relevant to mixing and coherence in real space in the same sense in which the Lagrangian description of a two-dimensional fluid defined in physical space by the fluid velocity field obtainable from a stream-function can be mapped on to a Hamiltonian flow in two dimensional phase space [23, 5]. The evolution of the trajectories of the system under the Lagrangian description, corresponds to the evolution of tracer particles in the fluid. The system is effectively many-dimensional as many initial conditions evolve at the same time. Such a picture can be extended even to turbulent, spatially extended problems, such as turbulent jets [19]. Control methods for such systems must be capable of constraining or mixing many system trajectories. It will be seen that the control methods under discussion succeed in achieving this objective.

We review a variety of control methods which address some of the issues raised above. We discuss methods for the enhancement of chaos, mixing and mixing rates in § 2. We discuss methods which work towards enhancing mixing via enhancing stretching rates, via the addition of noise and by preventing the formation of islands in time periodic Hamiltonian flows. Section 3 discusses methods for the formation of coherent structures in chaotic flows and maps. The methods discussed include that of introducing a fold and that of using localized parameter perturbations in the vicinity of low order periodic points. Section 4 discusses phenomena involving active tracers viz. aggregation and breakup. A brief discussion (§ 5) concludes this review.

2. Enhancing chaos, mixing and the rate of mixing

Enhancing mixing and the rate of mixing in a process is a topic of much recent research, for the reasons of obvious practical advantage listed above. Recently some novel methods to achieve this goal have been proposed [22, 24, 25]. These processes use different ways to increase mixing, and the rate of mixing. While one of the proposed methods works for time periodic flows by exploiting the underlying symmetries of the flow to prevent the island formation which constitutes a barrier to efficient mixing [22], another method [25] considers the effect of noise on such flows and its role in enhancing mixing. Recently, a new method has been proposed [24], which enhances the rate of mixing of chaotic flows by enhancing their chaoticity. This method enhances the chaoticity as measured by the average Lyapunov exponent by optimally perturbing the system parameters in regions of low local Lyapunov exponent. This increase in chaoticity translates into an enhancement of the rate of mixing, and can also be useful in other contexts where the enhancement and maintenance of chaos is important [4]. We summarize this method below.

2.1 *Enhancing chaos and the rate of mixing via control of stretching rates*

This method has the two-fold objective of enhancing chaos and thereby enhancing the rate of mixing of the system. An important parameter, which characterizes the degree of chaos in a chaotic flow is the Lyapunov exponent, which gives the average rate of stretching. However, the rate of stretching is not uniform over a chaotic attractor in the case of dissipative flows or over the phase space of a conservative flow. Thus the local Lyapunov exponent (LLE), a measure of the local rate of stretching, is different in different regions of the phase space. The nonuniform nature of the spatial distribution of the LLEs is exploited to construct a mechanism that can enhance chaos and hence the rate of chaotic mixing. The average rate of stretching is enhanced by introducing a small parameter perturbation which enhances the LLE whenever the system trajectory visits a region where the LLEs take values much smaller than their average value. This procedure works quite efficiently as small perturbations in parameter made for small times compared to the total time of evolution can lead to substantial enhancement of the Lyapunov exponent and to the rate of mixing. The method is applicable to both dissipative and conservative examples and discusses mixing in phase space. However, here as in all other contexts considered in this review, the conservative examples are relevant to mixing in real space in the following sense. Consider a two-dimensional fluid with a fluid velocity field $(u(x,y,t), v(x,y,t))$. For an incompressible fluid, we have

$$u_x + v_y = 0. \tag{1}$$

The fluid velocity field can be obtained from an exact differential $\mathrm{d}\psi$ using the equations

$$u = \frac{\partial\psi}{\partial y}, \quad v = -\frac{\partial\psi}{\partial x}, \tag{2}$$

where $\psi = \psi(x,y,t)$ is known as the stream-function.

The Lagrangian description in two dimensions involves following fluid particle trajectories (e.g. those of tracer particles) obtained by solving the equations

$$\dot{x} = u(x,y,t), \quad \dot{y} = v(x,y,t). \tag{3}$$

Using equation 2 above, we have

$$\dot{x} = \frac{\partial\psi(x,y,t)}{\partial y}, \quad \dot{y} = -\frac{\partial\psi(x,y,t)}{\partial x}. \tag{4}$$

Thus the stream-function behaves like the Hamiltonian of a Hamiltonian flow in two-dimensional phase space. The evolution of tracer particle trajectories in the two-dimensional fluid has been mapped on to the Hamiltonian dynamics in two-dimensional phase space. Thus, statements about the mixing or coherence of the phase-space trajectories map on to the behaviour of the tracer particle trajectories.

Consider an autonomous nonlinear dynamical system, of dimension n, evolving via the equations

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \mu), \tag{5}$$

where the set of parameters μ takes values such that the trajectory shows chaotic behaviour. Let $\mathbf{w}(\mathbf{x}, t)$ be the tangent vector to the trajectory at the point $\mathbf{x}$ and time t. The evolution of $\mathbf{w}$ is given by

$$\dot{\mathbf{w}} = (\mathbf{w} \cdot \nabla)\mathbf{F}. \tag{6}$$

The Lyapunov exponent of the system is defined by

$$\lambda = \lim_{t\to\infty} \frac{1}{t} \ln \frac{||\mathbf{w}(\mathbf{x}, t)||}{||\mathbf{w}(\mathbf{x}(0), 0)||}, \tag{7}$$

where $\mathbf{x}(0)$ is the value of $\mathbf{x}$ at $t = 0$ and $||\mathbf{w}||$ is the norm of $\mathbf{w}$. We now define the local Lyapunov exponent $\lambda(\mathbf{x})$ as

$$\lambda(\mathbf{x}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \ln \frac{||\mathbf{w}(\mathbf{x}(t+\Delta t), t+\Delta t)||}{||\mathbf{w}(\mathbf{x}(t), t)||} . \tag{8}$$

The quantity $\lambda(\mathbf{x})$ represents the local rate of stretching at the point $\mathbf{x}$. This is, in general, not uniform over the attractor. Clearly, the Lyapunov exponent λ (eq. (7)) is the average value of the LLEs for a long orbit .

The control mechanism is activated in the regions where the local Lyapunov exponent falls below its average value, that is if

$$\lambda(\mathbf{x}) < (\lambda - \gamma\sigma_\lambda) \tag{9}$$

where σ_λ is the standard deviation of the distribution of local Lyapunov exponents and γ is some chosen factor. The control is activated so as to change the parameter μ of the system to $\mu + s\,\mathrm{d}\mu$. Here s takes values -1 or $+1$ and $\mathrm{d}\mu$ is small. The control remains active as long as the above condition is satisfied, after which it is reset to the previous value.

The sign of s is decided in the following manner. Let W^T be a row vector and M^T given by $M^T = \nabla \mathbf{F}$. The equation for the norm of W can be written as

$$||\dot{W}||^2 = W^T(M^T + M)W. \tag{10}$$

The rate of change in the norm due to the change in the parameter should be positive in order to increase the local rate of stretching i.e. s should be chosen such that the quantity

$$\Delta||\dot{W}||^2 = ||\dot{W}(\mu + \mathrm{d}\mu)||^2 - ||\dot{W}(\mu)||^2 \simeq W^T({M_\mu}^T + M_\mu)W\mathrm{d}\mu \tag{11}$$

where $M_\mu = \partial M/\partial\mu$, should be positive.

The above equation is obtained from an expansion to linear order, however the effect of the change in parameter is nonlinear and often changes in the Lyapunov exponent are much greater than those predicted by a linear equation like (11).

The above control procedure gives rise to a substantial enhancement of the chaoticity of the system, as measured by the average Lyapunov exponent. This can be demonstrated by plotting the Lyapunov exponent for both controlled and uncontrolled flows. This is demonstrated in figure 1 for the Lorenz system [26] at parameter values $\sigma = 10.0, r = 30.0$ and $b = 8/3$. The parameter r is chosen as the control parameter and $\mathrm{d}r = 1.0$, $\gamma = 0.5$. The Lyapunov exponent is enhanced from 0.95 to 1.44. Figure 1 shows the behaviour of λ with r in both controlled and uncontrolled systems. As can be clearly seen, there is a substantial enhancement in λ over a range of parameter values.

It can also be demonstrated that this change in the Lyapunov exponent actually gives rise to an enhancement of the rate of mixing. The attractor is covered with a grid of boxes and a large number of initial conditions are chosen in a particular box. These initial conditions are then allowed to evolve without the control. The same set of initial conditions is allowed to evolve separately with the control procedure operative. The initial conditions in one box spread over a number of boxes in course of evolution. A comparison of the number of occupied boxes, i.e. the number of boxes which have at least one of the evolving initial conditions as a function of time for the controlled and

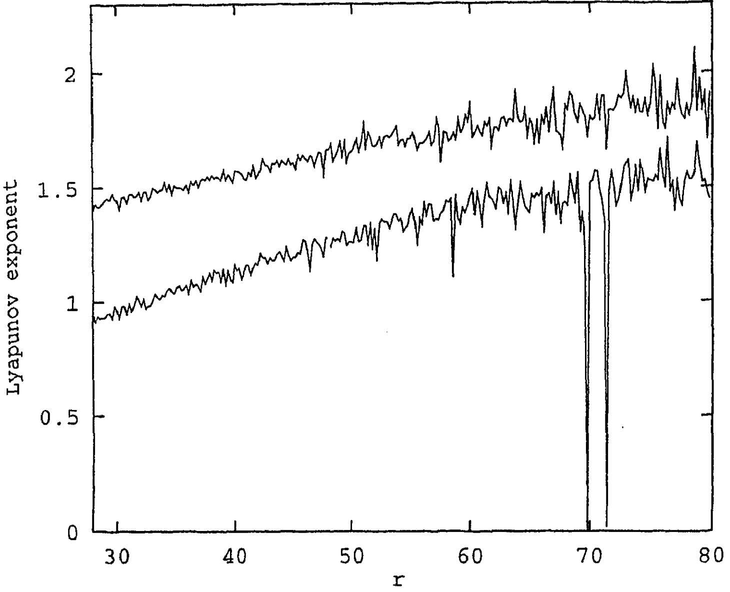


Figure 1. The Lyapunov exponent for Lorenz system at parameter values $\sigma = 10.0$, $r = 30.0$ and $b = 8/3$. The parameter r is chosen as the control parameter. The figure shows the behaviour of λ with r in both controlled and uncontrolled Lorenz system.

uncontrolled systems gives an idea of the relative rates of mixing of the two systems. Figure 2 plots this number for both controlled and uncontrolled cases for Lorenz system. It is clear that the controlled system mixes at a faster rate than the uncontrolled one, and that the control procedure has successfully manipulated the entire ensemble of trajectories in a manner in which the rate of mixing is enhanced.

The procedure used above to enhance chaos and the mixing rate can be easily modified to apply to the case of discrete maps. For maps, the evolution equation (eq. (5)) can be written as

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \mu), \tag{12}$$

where $\mathbf{x}_t$ are the dynamical variables at time t. The evolution of the tangent vector $\mathbf{w}$ is given by

$$\mathbf{w}_{t+1} = (\mathbf{w}_t \cdot \nabla)\, \mathbf{f}. \tag{13}$$

The control procedure is the same as above. The parameter μ is changed to $\mu + s\mathrm{d}\mu$ when condition (9) is satisfied. To decide the sign s we write (13) in matrix form as

$$W_{t+1} = MW_t, \tag{14}$$

where $M^T = \nabla \mathbf{f}$. The equation for the norm of W is

$$||W_{t+1}||^2 = W_t^T M^T M W_t. \tag{15}$$

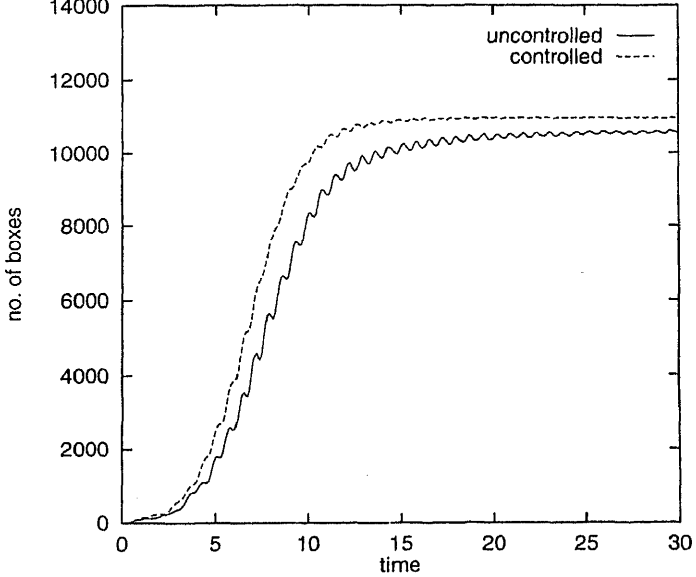


Figure 2. The number of occupied boxes as a function of time for the controlled and uncontrolled Lorenz systems.

Thus the rate of change in the norm of W due to change in the parameter is given by

$$\begin{aligned}\Delta||W_{t+1}||^2 &= ||W_{t+1}(\mu+\mathrm{d}\mu)||^2 - ||W_{t+1}(\mu)||^2 \quad (16)\\ &= W_t^T(M^TM_\mu + M_\mu^TM)W_t\,\mathrm{d}\mu \quad (17)\end{aligned}$$

where the last step is obtained by expanding to lowest order in $\mathrm{d}\mu$ and $M_\mu = \partial M/\partial\mu$. For control to enhance chaos and rate of mixing, the sign s for the parameter change $\mathrm{d}\mu$ must be such that $\Delta||W_{t+1}||^2$ is positive.

The Lorenz system discussed above is an example of a dissipative flow. However, the procedure is completely general and can be applied to conservative systems as well, provided the change in parameter is made in a way in which the conservative nature of the flow is preserved. We demonstrate the increase in chaos and rate of mixing in the standard map [23].

$$\begin{aligned}X_{t+1} &= X_t + \frac{K}{2\pi}\sin(2\pi Y_t)\\ Y_{t+1} &= Y_t + X_{t+1} \quad (18)\end{aligned}$$

where the variable X and Y lie between $[0,1]$. The increase in the rate of spread of initial conditions in phase space of the controlled standard map as compared to the uncontrolled one is demonstrated in figure 3. The parameter values are $K = 1.5, \mathrm{d}K = 0.2, \gamma = 0.25$. It can be easily seen that the controlled map retains its area-preserving property. We iterate $10,000$ initial conditions initially in a small box. Figures 3(a) and 3(b) show the uncontrolled and controlled systems respectively after 10 iterations. Figures 3(c) and 3(d)

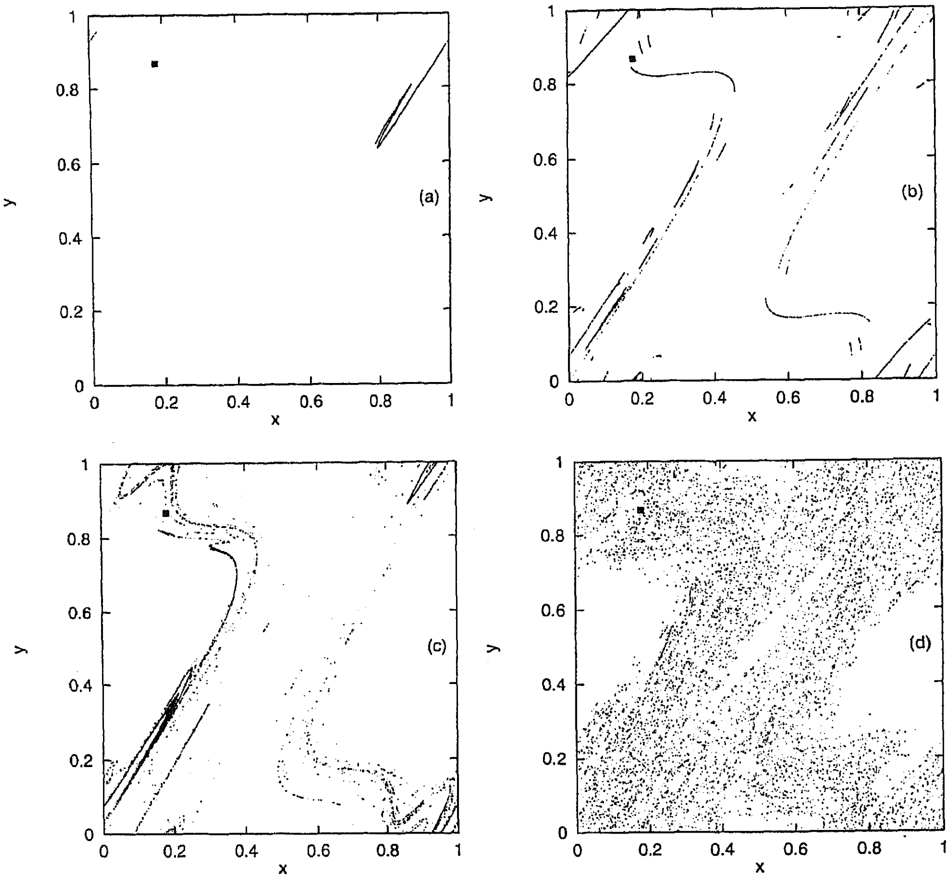


Figure 3. (**a**) and (**b**) show the uncontrolled and controlled systems respectively after 10 iterations. (**c**) and (**d**) show the corresponding figures after 30 iterations.

show the corresponding figures after 30 iterations. It is clear that the controlled system has spread out more than the uncontrolled situations. Thus the rate of mixing is enhanced. The Lyapunov exponent for the orbit with the initial condition $x = 0.2, y = 0.3$ increases from $\lambda = 0.286$ to $\lambda = 0.384$. The invariant KAM tori in the phase space of the standard map function as barriers to global mixing. If the effect of control on asymptotic global mixing is observed, the controlled and uncontrolled attractors look very similar in both cases except that the area of empty regions in the uncontrolled situation is larger than the area of empty regions in the controlled situation.

The control procedure works quite efficiently as it produces a substantial enhancement of the Lyapunov exponent for quite small changes in the parameters. This is due to the fact that the control procedure works by switching between three types of chaotic flows, those characteristic of parameter values μ, $\mu + \mathrm{d}\mu$ and $\mu - \mathrm{d}\mu$. This switching introduces an extra time dependence in the problem and is the origin of the efficiency of the procedure.

Thus, this mechanism is successful in enhancing the rate of mixing of the system and also in enhancing the chaoticity of the system. It can therefore be useful in contexts where the maintaining and enhancement of chaos is important. Other methods of enhancing mixing have also been proposed in various specific contexts. A very simple way of enhancing the mixing in a flow is by the addition of noise.

2.2 *Noise and mixing*

The Lagrangian trajectory of particle is given by (3). However in realistic situations there is a stochastic correction applied to this equation. This correction may be due to the Brownian motion of the constituent particles or due to extrinsic noise.

The equation of motion for a diffusive particle is given by the generalized Langevin equation where

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{u}(\mathbf{x}, t) + \zeta(t) \tag{19}$$

and $\mathbf{u}(\mathbf{x}, t)$ is the velocity at point $\mathbf{x}$ at time t. The probability density function of the stochastic term, $\zeta(t)$ is a gaussian and the variance of this distribution is related to the molecular diffusivity by $\sigma^2 = 2D\Delta t$, where Δt is the time interval between displacements. To apply this kind of term to discrete maps, the particles are stochastically displaced after each iteration of the map. The addition of such a noise term enhances mixing. This has been explicitly demonstrated [25] in the case of the blinking vortex map [18].

Another method which enhances global mixing has been proposed for time periodic flows.

2.3 *Symmetries and island formation*

Time periodic flows can be modelled by a Hamiltonian system with the streamfunction acting as the Hamiltonian [23]. When a KAM curve for such a system is perturbed, it breaks into elliptical and hyperbolic points. Islands are formed around elliptic points, which are regions of regular behaviour separated by an invariant curve from the chaotic component. These islands affect the transport and mixing in two ways: firstly they do not exchange trajectories with the chaotic counterpart and hence act as trapping regions, and secondly, due to their 'stickiness', trajectories which come close to these islands spend a larger fraction of time in the vicinity. Both these properties are a hindrance to efficient mixing. Therefore, to increase the mixing, one should control the flow in such a way so as to prevent the formation of these regions.

It has also been observed that in many cases, the placement of these periodic points in the phase space is symmetric [27]. It is possible to use these symmetries to locate these periodic points, and to devise a protocol which would prevent the formation of these islands [22].

An interesting demonstration of this method can be made in the case of the cavity flow, where the system consists of fluid in a cavity whose two walls are moving transversally (say the top and bottom walls), whereas the other two walls are held rigid. The motion is a composition of two independent operations, that due to the top wall $\mathbf{T}_t$ [this means that a set of particles $\mathbf{x}$ is mapped to a set $\mathbf{T}_t\mathbf{x}$ after time t under the action of the top wall alone] and that due to the bottom wall $\mathbf{B}_t$. The system is completely characterized by a sequence of top and bottom wall motions. If we have a periodic operation of $\mathbf{T}_t\,\mathbf{B}_t$ each operating for equal duration of time then the system has x-axis as the symmetry curve.

To enhance mixing one tries to use the symmetries in such a way that the islands are never formed . The flow **BTBTBT**... mentioned above has x symmetry and the islands lie on the x-axis or in pairs symmetrically distributed on both ides of x-axis. One way to prevent the formation is to follow the flow $\mathbf{Q}_0 = \mathbf{BT}$ by an identical flow which is rotated by 180°. That is, now the flow is $[\mathbf{R}(\mathbf{BT})\mathbf{R}]\mathbf{BT} = \mathbf{Q}_1 = \mathbf{TBBT}$. This protocol switches the position of the islands such that they never get a chance to form. Other protocols which prevent island formation are also possible, e.g. a switch in the flow after two periods will also give similar results. These protocols have been tested experimentally for the cavity flow [22] and indeed increase the global mixing by removing the islands.

Thus, by exploiting the underlying symmetries of the flow, one can remove the islands. Since the islands hamper mixing by both trapping as well as by slowing down the trajectories, this removal enhances mixing. However, as long as the sequence of protocol remains finite, there will always be islands present in the flow which will not be removed by this sequence. Secondly, these protocols can be applied only in case of time periodic, incompressible, nondissipative flows. The applicability of such protocols in case of flows that are dissipative or non-periodic remains unclear.

3. Coherent structures

As mentioned earlier, the existence of localized, coherent structures in flows that are otherwise chaotic or turbulent is a commonly observed feature of natural systems [14, 13]. Again, the reasons for the formation of coherent structures in otherwise irregular flows can be investigated using the well known fact that spatially extensive systems can be modelled by low dimensional chaotic flows [17]. Recent work suggests that control methods which can create coherent structures in chaotic flows and maps may provide useful insights into the ubiquitous existence of these objects in natural systems. We outline two such methods.

The first method creates the coherent structure by the introduction of a fold in the vicinity of the periodic points of the system [28]. The second method uses localized parameter perturbations in the neighbourhood of the fixed points to create the coherent structures [29].

3.1 *Folding*

Chaotic flows are typically flows which stretch and fold. Such flows are usually well mixing. Thus there is an intimate relation between stretching and folding and the mixing properties [30]. The prototype for describing the stretching and folding of flows is a horseshoe construction. These horseshoes are formed in chaotic regions of the flow. A

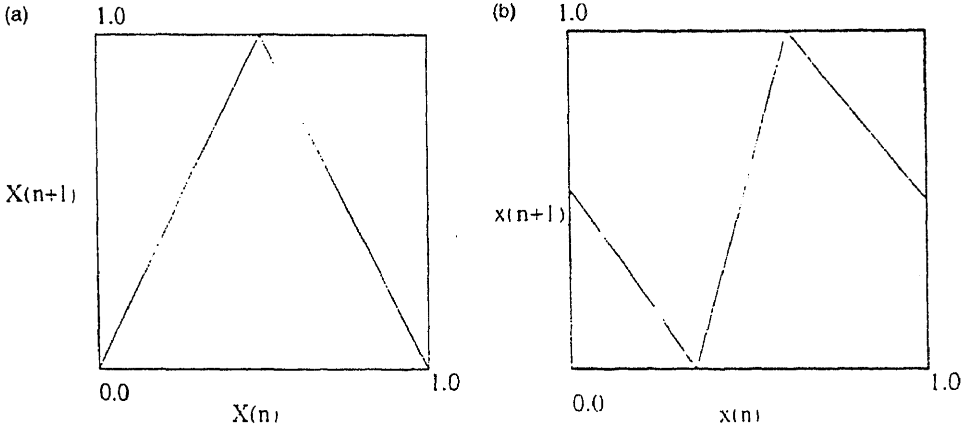


Figure 4. (**a**) The tent map in the unit interval [0, 1]. (**b**) shows the tent map with an additional fold introduced. This additional fold is responsible for trapping of trajectories.

horseshoe is termed *complete* if the fold is located outside the initial square and *incomplete* if the fold is located inside. Complete horseshoes generate efficient mixing where as incomplete horseshoes do not. Shinbrot and Ottino [28] show that the proximity of the fold of an underlying horseshoe map to a periodic point characterizing the order of the horseshoe can produce an incomplete horseshoe and generate an invariant region or a coherent structure.

This is easiest to illustrate in the case of the tent map (figure 4(a)) given by

$$\begin{aligned} X_{n+1} &= 2X_n \qquad 0 \le X_n \le \tfrac{1}{2} \\ &= 2(1-X_n) \qquad \tfrac{1}{2} \le X_n \le 1. \end{aligned} \tag{20}$$

Every small interval between 0 and 1 is stretched by a factor of 2 after every iteration of the map (20). It is therefore impossible for any such interval to contain a localized structure. The only exception to this occurs near a fold where the interval is doubled over, thus conceivably counteracting the stretching. To construct an invariant interval I by using folds, a fold must be placed sufficiently near a fixed point that points within I remain within I after successive iterations of the map.

In the case of the tent map, it is easy to achieve this by changing the phase, α, as defined in the new map

$$\begin{aligned} X_{n+1} &= 2(\alpha - X_n) \quad 0 \le X_n \le \alpha \\ &= 2(X_n - \alpha) \quad \alpha \le X_n \le \tfrac{1}{2} + \alpha \\ &= 2(1+\alpha-X_n) \quad \tfrac{1}{2}+\alpha \le X_n \le 1. \end{aligned} \tag{21}$$

This map (figure 4(b)) now contains both a fixed point at $X_{n+1} = X_n = (2\alpha/3)$ and a fold at $X_n = \alpha$ in the interval $I = (0, 2\alpha)$ and points initially in the interval I remain trapped forever inside the interval I despite repeated iterations of the map (21). The effect of the fold is to create a trapping region in the map. In the case of the tent map with the fold no new coherent structure is created despite the trapping of trajectories.

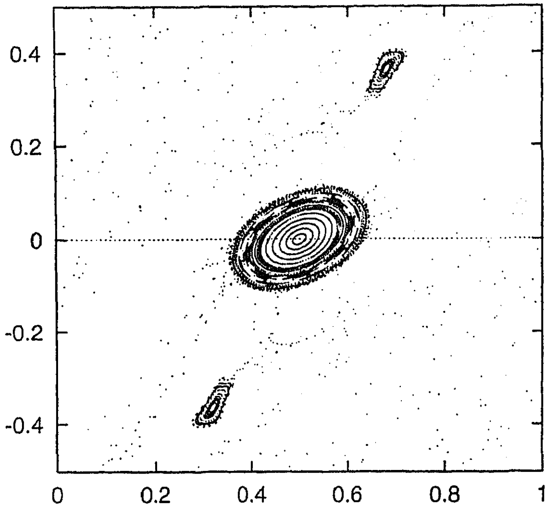


Figure 5. This figure shows the phase space plot of a standard map in presence of an additional fold. The map parameter $K = 5.1$ and the fold is introduced in the region $0.3 < x < 0.7$ as can be clearly seen this results in a coherent structure being formed in the neighbourhood of the fixed point at $(0.5, 0.0)$. The values of the constants in the cubic polynomial are $a_0 = 0.0069$, $a_1 = -0.713$ and $a_2 = 1.0023$, $a_3 = 0.321$.

The effect of the fold in creating a coherent structure can be easily demonstrated in the context of the standard map (18). The variable X is restricted to [0, 1] and the variable Y to $[-0.5, 0.5]$. The map shows chaotic behaviour for parameter values exceeding 0.975 and has a hyperbolic fixed point at $(0.5, 0)$. A fold is introduced in the region R around the fixed point in the following manner. We replace $(K/(2\pi))\sin(2\pi X_n)$ by a functiom $F(X_n)$ when (X_n, Y_n) lies in R. To produce a fold $F(X)$ is chosen so as to contain at least one extremum in R with $F(X) = (K/(2\pi))\sin(2\pi x_n)$ at the bounds of R. A suitable choice is the modified cubic polynomial $F(X) = a_0 + a_1(X - \frac{1}{2}) + a_2(X - \frac{1}{2})^2 + a_3(X - \frac{1}{2})^3$ where a_0, a_1 and a_2 are chosen so as to match $F(x)$ with $(K/(2\pi))$ $\sin(2\pi x_n)$ at $X = 0.3$ and $Y = -0.2$. It is clear that the effect of the fold is to constrain a finite fraction of the ensemble of trajectories to an invariant neighbourhood of the fixed point in spite of the presence of stretching. The resulting coherent structure is shown in figure 5.

The same idea can be applied to experimentally realizable systems and has been implemented by Shinbrot and Ottino for the numerical realization of the journal bearing flow where a fluid constrained between two asymmetrically placed cylinders is chaotically advected by driving the inner and outer cylinders alternately [23, 6]. Control is implemented by controlling the clockwise and anticlockwise rotation of the outer and inner cylinder. The tracer particles of the fluid are chaotically advected for a repeating protocol which drives the inner cylinder clockwise through 1080°, outer anticlockwise through 240° and outer through 1080°. A fold leading to a coherent structure in this case can be provided by a sequence which drives the outer cylinder clockwise through 32°, followed by rotations of the inner cylinder through 1080°, the outer anticlockwise through 170° and the inner through 1080° [28].

The robustness of these structures has been studied by adding a noise term. The structure is stable for small noise strengths but is destroyed for large noise. An interesting application of these coherent structures is to use them as controlled release capsules where one or more reactants of a chemical reaction can be confined in a coherent structure while the remaining flow is mixed. At a desired moment the coherent structure can be destroyed by the application of a perturbation thus releasing the reactant in the chemical reaction. This can be used to control various reaction rates and products [20].

The fold, however varies from case to case and does not have a generic form. Also the placement of such a fold could be difficult in various realistic situation and time consuming. A new method to produce these structures via parameter perturbations has been proposed in [29]. This method is natural and generic and is also easier to implement than the method of the fold.

3.2 *Localized parameter perturbation*

We now summarize a new method by which coherent structures can be created in two-dimensional area-preserving maps [29]. Such maps can serve as representations of realistic fluid flows [18, 31]. Our method consists of the application of parameter perturbations in the local neighbourhoods of the periodic points of the system. The perturbation is applied in a fashion in which the area-preserving nature of the map is retained.

It is easy to see that this control mechanism is a natural mechanism in the context of realistic flows where parameter perturbations in localized neighbourhoods can arise in a variety of situations in nature and in experiment. Examples of these are stirring protocols in experiment and convective flows in environments of non-uniform temperatures. Noise can also be a source of such perturbations. The localized variation of parameters such as density are believed to be responsible for the formation of coherent vortices in planetary and oceanic contexts [13, 32]. Experiments designed to simulate Jovian and other planetary surfaces have successfully created vortices by incorporating localized perturbations [33].

The simplest example where this method can be demonstrated is the two-dimensional area-preserving standard map given by (18). As said before, the map shows chaotic behaviour for parameter values exceeding 0.975 and has a hyperbolic fixed point at $(0.5, 0)$. The system is controlled in the following manner. The parameter K in (18) is replaced by the parameter k where $k = K - \epsilon$ if $|X - X_f| < \delta, |Y - Y_f| < \delta$ and $k = K$ for values of x, y which lie outside this square. Here (X_f, Y_f) is a period-1 point of the uncontrolled map (18) and ϵ is a control parameter. It is clear that the dynamics of the controlled map is area-preserving as the Jacobian $\mathbf{J}$ remains unity.

Figure 6(a) shows the phase-space plot of the uncontrolled map (18) for $K = 6.0$ with 200 initial conditions chosen randomly over the unit square, iterated 500 times. It is clear that the map is completely chaotic and the distribution of points is uniform over the phase space. Compare this with the phase-space plot (figure 6(b)) of the controlled map described above where the fixed point (x_f, y_f) is chosen to be the hyperbolic fixed point $(0.5, 0.0)$, and $\delta = 0.1$, $\epsilon = 2.0$ iterated the same number of times for the same initial conditions. It is clear that a prominent new coherent structure now appears in the neighbourhood of the fixed point which was absent in the original standard map. Thus

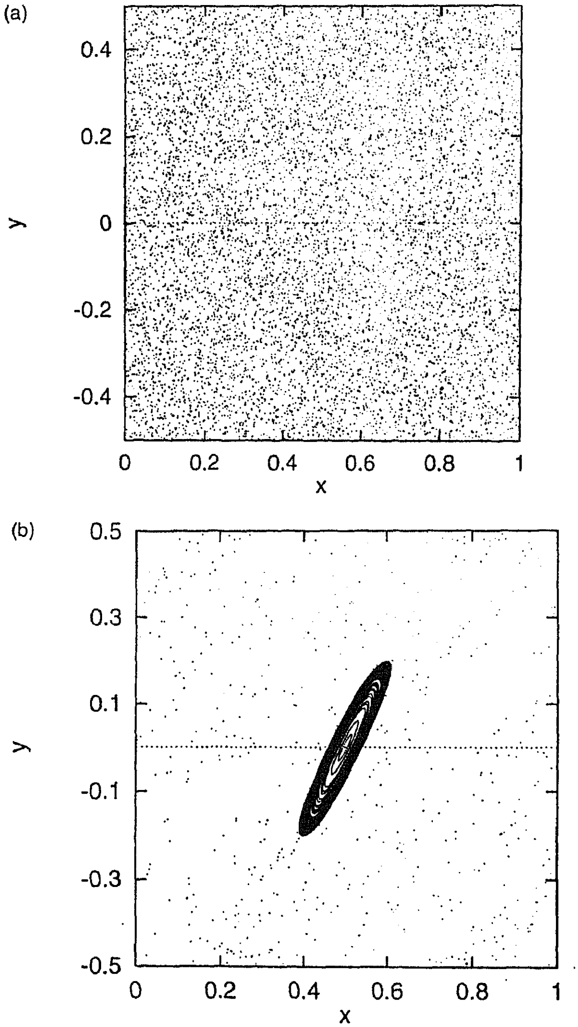


Figure 6. (**a**) The phase-space plot for the uncontrolled standard map. (**b**) The formation of the coherent structure. The values of K, ϵ, δ are $6.0, 2.0$ and 0.1 respectively, with $(x_f, y_f) = (0.5, 0.0)$.

perturbing the map parameter in the vicinity of the fixed point constrains a part of the ensemble of trajectories to give new coherent structures. The method can also be successfully demonstrated in case of maps representing realistic fluid situations, e.g. in the Blinking vortex map [18].

If the local rates of stretching i.e. local rates of divergence of trajectories in different regions of phase space are calculated, the controlled system shows decreased rates of divergence (as compared with the uncontrolled system) in the regions where coherent structures are formed [29]. The decreased divergence of the trajectories has the effect of consequence. The reduction in local mixing tends to favour coherence. It is interesting to compare this method with the control procedure of § 2.1. The control there works by tailoring perturbation towards increasing the mixing tendency by increasing the local Lyapunov exponents in regions of low local Lyapunov exponent. This increases the diverging tendency of trajectories and hence the mixing. Conversely, the present method works by decreasing the divergence rate of trajectories in the neighbourhood of low-order periodic points, another way of locating localized neighbourhoods where the divergence rates are of significance.

A comparison of the present method with the introduction of fold discussed in previous section is also interesting. In contrast to the introduction of a fold in the vicinity of the periodic point the present method introduces a parameter perturbation which modifies the distribution of the stretching rates. Since the perturbation retains the area-preserving nature of the systems, the distribution of contraction rates (the distribution of the smaller Lyapunov exponents of the system) correspondingly compensates for the alteration. This in effect modifies the folding of the system without the explicit introduction of an additional fold. In both cases, the control of the flow in the neighborhood of low order periodic points produces coherence, but the nature of the periodic points (whether hyperbolic or elliptic) does not appear to be of significance.

Thus we see that coherent structures can be created in simple chaotic flows using some simple control techniques. These structures can be creatively used for various applications and the control mechanism described can provide a useful insight into the ubiquitous presence of these structures in various situations as well as the mechanism responsible for their creation. The structures can be characterized by using some local quantifiers. Such an approach may also shed light on the way these structures affect various transport quantities in chaotic flows.

The flow region thus has a variety of objects, ranging from fast mixing chaotic regions to islands and coherent structures. This backdrop provides an underlying template for various interesting processes like aggregation, coagulation and breakup etc. The presence of efficient mixing or coherent structures modifies and greatly influences these processes. Therefore, by controlling various mixing properties or by creating coherent structures, one may be able to greatly influence these processes to one's advantage.

4. Mixing, aggregation and breakup

There are a lot of interesting phenomena which are associated with mixing in fluid flows. These phenomena deal with situations where the particles in fluid flow are no longer passive but interact with each other. They collide, stick together, aggregate. A fluid element elongates, stretches and breaks up. In other words, in a background of chaotic flows revealed by passive tracers, a number of phenomena involving active tracers can take place.

One of the most common phenomena which occur in the case of active tracers is the break up of droplets. The crucial factors in this context are the velocity field of the fluid, the viscosity of the droplet and surface tension effects [34]. The breakup in case of chaotic flows depends on the placement of the droplet. If the droplet is placed in an island then it undergoes a periodic sequence of stretching and contraction and there is no net stretching. If the viscosity is strong enough to overcome surface tension effects then the droplets placed in the chaotic region become elongated into long filaments. This filament is then stretched, compressed and folded in a complicated way. The breakup of these filaments can occur by various mechanisms such as capillary wave instabilities, endpinching, necking etc. However there are some modes of breakup which are peculiar to the nonlinear chaotic flows. The prime example is the breakup in folds, which is associated with change in sign of stretching rate [6]. In most of these modes of breakup, the breakup is directly dependent on the mixing properties of the flow. The repetitive nature of the mixing involving a number of stretching and folding generates an equilibrium drop size distribution which is self similiar. The crucial factor in the breakup is the stretching rate. Clearly, the control methods discussed above can be used to redistribute the stretching rates and to modify the rate of mixing so that the break up of droplets is accelerated or slowed down depending on the application.

The second interesting process involving active tracers is the process of aggregation. This is the reverse process of breakup. The simplest scenario involving aggregation or coagulation is that of particles being advected by a flow and interacting with each other over short range interactions. For example, if two particles come together within a capture diameter d, they stick to each other and move together subsequently. Thus the size of the particle clusters grows till it becomes too big when mixing takes over and breaks it up thus limiting further growth of these clusters. Hence once again, the size to which such a cluster would grow is dictated by localised mixing properties. Coherent structures, with their trapped and localized trajectories could act as preferred sites where large structures could grow. Thus understanding and controlling these localized properties is once again the key to create tailor made aggregates of desired size in various flows. The case of the growing cluster is also reminiscent of those formed by diffusion limited aggregation. The aggregates can be isotropic or highly anisotropic depending on the presence of islands and other coherent structures. Control methods could be used to promote desired cluster configurations.

All the interactions described here are 'one-way'. The way the particles interact is influenced by the flow and various local mixing properties. However, the presence of such interaction may also change the nature of flow and hence influence the whole process. For example, the addition of small amounts of polymer to a Newtonian fluid decreases the amount of chaos in the flow [35]. The underlying presence of chaos, however still persists. These phenomena are little investigated up to now and much remains to be understood. Again, control methods may provide insights into these problems.

5. Conclusion

We have summarized some recently proposed applications of control methods to problems of mixing and coherence in chaotic dynamical systems. This is an important problem both for its own intrinsic interest and also from the point of view of applications. These methods provide insights into the origin of mixing and unmixing behaviour in natural systems. They could be useful for improving the rate of mixing of flows, enhancing global mixing, and also for creating coherent regions in flows. The results could find applicability to fluid flows, chemical reactions, oscillator arrays and other extended systems and are relevant for promoting efficient engineering processes and tailoring catalytic reactions. One of the methods discussed is also useful for enhancing the chaoticity of dynamical systems and could be useful in biomedical applications where the destruction or reduction of chaos could have pathological consequences. These methods could also be used for tailoring aggregation and break-up processes for suspended particles in fluid flows. We hope to explore some of these directions in future.

References

[1] E Ott, C Grebogi and J A Yorke, *Phys. Rev. Lett.* **64**, 1196 (1990)
L M Pecora and T L Caroll, *Phys. Rev. Lett.* **64**, 821 (1990); *Phys. Rev.* **A44**, 2374 (1990)
[2] M A Woo, W G Stevenson, D K Moser, R M Harper and R Trelease, *Am. Heart J.* **123**, 704 (1992)
A Goldberger, *Ann. Biomed. Eng.* **18**, 195 (1990)
[3] S J Schiff, K Jenger, D H Duong, T Chang, M L Spano and W L Ditto, *Nature* **370**, 615 (1994)
[4] W Yang, M Ding, A Mandell and E Ott, *Phys. Rev.* **E51**, 102 (1995)
[5] Proceedings of an International Symposium on Fluid Mechanics of Stirring and Mixing, *Phys. Fluids* **A3** (1991)
J M Ottino, *Sci. Am.* **260**, 40 (1989)
J M Ottino, F J Muzzio, M Tjahjadi, J G Franjione, S C Jana and H A Kusch, *Science* **253**, 755 (1995)
J M Ottino, G Metcalfe and S C Jana, in *Proceedings of the Second Experimental Chaos Conference* (World Scientific, Singapore, 1995) pp. 3–20
[6] J M Ottino, *The kinematics of mixing: stretching, chaos and transport* (Cambridge University Press, Cambridge, 1989)
[7] W R Hawthorne, D S Wendell and H C Holtell, Mixing and combustion in turbulent gas jets, in *Third symposium on combustion and flame and explosion phenomena* (Baltimore, Williams and Wikens, 1948)
[8] R S Spenser and R M Wiley, *J. Coll. Sci.* **6**, 133 (1951)
[9] K D Williamowski and O E Rossler, *Z. Naturforsch., Teil* **A35**, 317 (1980)
[10] B L Aguda and B L Clarke, *J. Chem. Phys.* **89**, 7428 (1988)
[11] A A Townsend, *Proc. R. Soc. London* **A209**, 418 (1951)
[12] P S Marcus and C Lee, *Chaos* **4**, 269 (1994)
[13] J S McWilliams and J B Weiss, *Chaos* **4**, 305 (1992)
[14] C W Leong and J M Ottino, *J. Fluid Mech.* **209**, 463 (1980)
[15] K Wiesenfeld and P Hadley, *Phys. Rev. Lett.* **62**, 1335 (1989)
[16] S D Meyers, J Sommeria and H L Swinney, *Physica* **D37**, 515 (1989)
[17] D S Broomhead, R Indik, A C Newell and D A Rand, *Nonlinearity* **4**, 159 (1991)
[18] H Aref, *J. Fluid Mech.* **143**, 1–21 (1984)
[19] H Aref, S W Jones, S Moffina and I Zawadzki, *Physica* **D37**, 423–440 (1989)
[20] T Shinbrot, G Metcalfe, L Bresler and J M Ottino, preprint
[21] F Muzzio, P D Swanson and J M Ottino, *Phys. Fluids* **4**, 471 (1991)
[22] J G Franjione, C W Leong and J M Ottino, *Phys. Fluids* **A1**, 1772 (1990)
[23] M Tabor, *Integrable and chaotic dynamical systems, an introduction* (J. Wiley, New York, 1990)
[24] N Gupte and R E Amritkar, *Phys. Rev.* (to be published)
[25] S W Jones, *Phys. Fluids* **A3**, 1081 (1991)
[26] E N Lorenz, *J. Atmos. Sci.* **20**, 130 (1963)
[27] C W Leong and J M Ottino, *Phys. Fluids* **A1**, 1441 (1989)
[28] T Shinbrot and J M Ottino, *Phys. Rev. Lett.* **71**, 843 (1993)
[29] Ashutosh Sharma and Neelima Gupte, unpublished
[30] W-L Chien, H Rising and J M Ottino, *J. Fluid. Mech.* **170**, 355 (1986)
[31] J Chaiken, C K Chu, M Tabor and Q M Tan, *Phys. Fluids* **30**, 687–694 (1987)

[32] J C McWilliams, *Rev. Geophys.* **23**, 165 (1985)
[33] S V Antipov, M V Nezlin, E N Snezkin and A S Trubnikov, *Nature* **323**, 238 (1990)
M V Nezlin, *Chaos* **4**, 187–202 (1994)
[34] B J Bentley and L G Leal, *J. Fluid. Mech.* **167**, 241 (1986)
H Stone, B J Bentley and L G Leal, *J. Fluid. Mech.* **173**, 131 (1986)
[35] J F Brady and G Bossis, *Ann. Rev. Fluid. Mech.* **20**, 111 (1989)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 1,
January 1997
pp. 249–258

Characterization and control of chaotic dynamics in a nerve conduction model equation

S RAJASEKAR
Department of Physics, Manonmaniam Sundaranar University, Tirunelveli 627 002, India

Abstract. In this paper we consider the Bonhoeffer–van der Pol (BVP) equation which describes propagation of nerve pulses in a neural membrane, and characterize the chaotic attractor at various bifurcations, and the probability distribution associated with weak and strong chaos. We illustrate control of chaos in the BVP equation by the Ott–Grebogi–Yorke method as well as through a periodic instantaneous burst.

Keywords. Bonhoeffer–van der Pol oscillator; local Lyapunov exponent; weak and strong chaos; controlling of chaos.

PACS Nos 05.45; 42.50

1. Introduction

During the last decade or so remarkable progress has been made in exploring the complexity of nonlinear systems, where the presence of chaos and other related phenomena have been extensively investigated [1, 2]. Particularly, routes to chaos, characterization of periodic and chaotic motions, multifractal, fractal basin boundary structures, etc. have been well established. In recent years there have been increasing interests in the study of

(i) statistical dynamics of local expansion rates [3–8],
(ii) characterization of chaotic attractors at bifurcations [9–14],
(iii) characterization of weak and strong chaos [15, 16],
(iv) controlling of chaos [2, 17–39] and
(v) synchronization of chaotic trajectories [2, 40–45].

1.1 *Local expansion rates and characterization of chaotic attractors at bifurcations*

Recently, the features associated with the fluctuations of local expansion rates of various attractors and statistical dynamics of local Lyapunov exponents have been studied [3–14]. Chaotic attractors have various local structures depending on their routes of onset and evolution. For instance, chaotic attractors are generally non-hyperbolic and have local homoclinic tangency structures. Bifurcations of chaos change the structures of chaotic and the variance $\sigma_n(q)$ of fluctuations of $\wedge_n(x_1)$ around $\wedge_n(q)$ are useful for describing the coherent large fluctuations at the bifurcations of chaos.

1.2 *Weak and strong chaos*

In the literature, chaotic motion with small positive Lyapunov exponent is called weak chaos and chaotic motion with sufficiently large positive Lyapunov exponent is termed as strong chaos. Recently, Chen and Chow [15] have shown that weak and strong chaos can be characterized using probability distribution of k-step difference of a variable of the given dynamical system.

1.3 *Controlling of chaos*

Even though the presence of chaotic behaviour is generic and robust there are practical situations where one wishes to avoid or control chaos so as to improve the performance of a dynamical system. The ability to control chaos that is to convert chaotic oscillations into a desired regular ones with a periodic time dependence, would be beneficial in working with a particular system. Recently, more interest has been focussed on controlling of chaos and various methodologies that control chaos [17–26, 33].

1.4 *Synchronization of chaotic trajectories*

An approach which is closely related to controlling of chaos is that of obtaining the phase space trajectory of a chaotic system synchronized with another chaotic trajectory [2, 40–45] which has practical applications in secure communication. Here, synchronization is obtained by setting some of the variables of the synchronizing system to the values of the desired trajectory. The remaining variables asymptotically synchronize with the desired values provided the subsystem Lyapunov exponents for these variables are negative.

Thus, there are many interesting topics in chaos theory which have received much attention during the last few years and many fascinating results have been found. We have carried out investigations in these directions with reference to certain discrete and continuous dynamical systems [8, 12–14, 16, 27–33]. The goal of the present paper is to study the salient features associated with some of the above aspects of chaos with reference to the Bonhoeffer–van der Pol (BVP) equation. The BVP oscillator is an interesting dynamical system of considerable physical and biological significance [46].

The plan of the paper is as follows. In § 2 we characterize the chaotic attractor at various bifurcations, using the variance $\sigma_n(q)$ of fluctuations of coarse-grained local expansion rates of nearby orbits. Next, the probability distribution $P(\Delta V_k)$, $k = 1, 2, \ldots$ of $\Delta V_k = V_{m+k} - V_m$ where $V_1, V_2, \ldots$ are the values of the variable V in the Poincaré map of the BVP system is investigated in § 3. A stationary $P(\Delta V_k)$ distribution is shown to occur for strong chaos while nonstationary $P(\Delta V_k)$ is observed for weak chaos. Then, in § 4 we consider the problem of controlling of chaos. We present a brief summary of previous results on controlling of chaos in the BVP system. We show the stabilization of unstable periodic orbit (UPO) of BVP oscillator by the method of Ott–Grebogi–Yorke [17], and the suppression of chaos by an additional instantaneous burst. Finally, § 5 contains summary and conclusions.

2. Characterization of chaotic attractors at bifurcations in the BVP oscillator

In this section we study the characterization of chaotic attractors at critical bifurcations with reference to the BVP equation [28, 47, 48] which describes the propagation of an electrical impulse or voltage pulse along the membranes of nerve cell:

$$\dot{V} = V - V^3/3 - R + A_0 + A_1 \cos t, \tag{1a}$$

$$\dot{R} = c(V + a - bR). \tag{1b}$$

The variable V can be seen as the electric potential across the cell membrane, while R is a so-called recovery variable in which dynamic aspects of the state of the membrane are aggregated. a, b, c are membrane radius, specific resistivity of the fluid and temperature factor respectively and $A_0 + A_1 \cos t$ is the membrane current.

For a chaotic orbit $\{X_n\}, n = 1, 2, \ldots$ generated by a two dimensional map, we define the coarse-grained expansion rates [9–11] as

$$\wedge_n(X_1) = (1/n) \sum_{m=1}^{n} \lambda(X_m), \tag{2}$$

where $\lambda(X_m)$ is the local expansion rate of nearby orbits at X_m along the unstable manifold. The values of $\wedge_n(X_1)$ for different X_1's are randomly distributed between minimum and maximum values. To describe the fluctuations of the local expansion rates we consider the partition function [9–11]

$$Z_n(q) = \langle \exp[-n(q-1) \wedge_n (X_1)] \rangle, (-\infty < q < \infty) \tag{3}$$

where $\langle \ldots \rangle$ denotes the long time average. The scaling exponents can be defined in terms of the partition function as

$$\phi_n(q) = -(1/n) \ln Z_n(q), \tag{4}$$

$$\wedge_n(q) = \mathrm{d}\phi_n(q)/\mathrm{d}q, \tag{5}$$

$$\sigma_n(q) = -\mathrm{d} \wedge_n (q)/\mathrm{d}q. \tag{6}$$

For the BVP oscillator $\lambda(X_m)$ along the unstable manifold at X_m is calculated in the Poincaré map. The dynamical structure functions $\wedge_n(q)$ and $\sigma_n(q)$ are calculated for a range of values of the control parameter A_1 about its critical values A_{1c} at which bifurcations such as sudden widening of a chaotic attractor, band merging and intermittency occur.

The parameters in (1) are fixed at $a = 0.7$, $b = 0.8$, $c = 0.1$, $A_0 = 0$. A band merging crisis is found to occur at $A_1 = A_{1m} \approx 1.28653$: as A_1 is decreased to A_{1m} three bands of a chaotic attractor merge to form a single band chaotic attractor. Figure 1a shows $\sigma_n(q)$ at $A_1 = 1.287$ far from A_{1m}, where we have used $n = 21$. The $\sigma_n(q)$ has a peak at $q = q_\alpha \approx 1.65$. $\sigma_n(q)$ just before the band merging is shown in figure 1b. The $\sigma_n(q)$ has two peaks at $q = q_\alpha \approx 2.4$ and $q = q_\beta \approx 0.4$. Two peaks are observed for A_1 values just

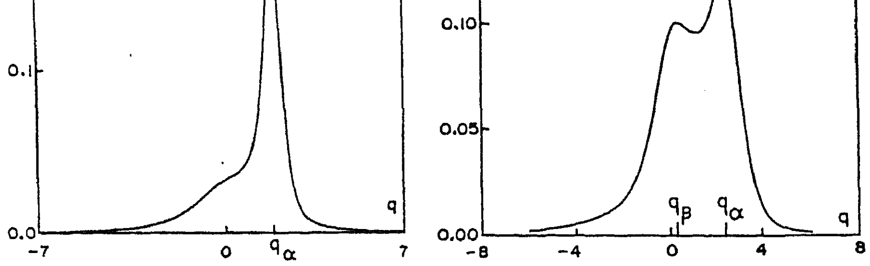


Figure 1. $\sigma_n(q)$ versus q for $A_1 = 1.287$ (**a**) far from band merging chaos and for $A_1 = 1.286$ (**b**) just before band merging chaos.

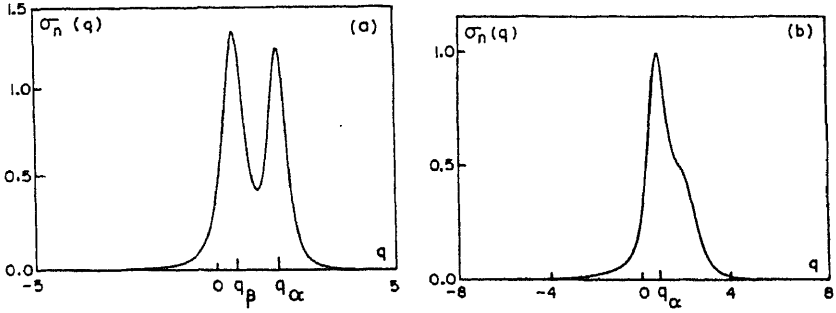


Figure 2. $\sigma_n(q)$ plots for (**a**) $A_1 = 1.0922$ in the intermittency region and (**b**) $A_1 = 1.096$ far after intermittency.

after the band merging also. For example, for $A_1 = 1.285$ peaks at $q_\alpha = 1.85$ and $q_\beta = -0.075$ are observed. Further, for $q = 1$, $\Lambda_n(1) = \Lambda^\infty$ and $\Lambda_n(\infty) = \Lambda_{\min}$ and $\Lambda_n(-\infty) = \Lambda_{\max}$. Therefore, $\sigma_n(q)$ with $q > 1$ and $q < 1$ can explicitly describe negative and positive large fluctuations of $\Lambda_n(X_1)$ respectively.

In the BVP equation, type I intermittent chaos is observed for $A_1 \in (1.092, 1.094)$. For A_1 values just below 1.092 a period-$4T (T = 2\pi)$ attractor is found. In the intermittency region the chaotic attractor has two types of local structures which produce the laminar and turbulent bursts respectively. Two types of local structures have been captured by the dynamical structure function $\sigma_n(q)$. Figure 2a shows $\sigma_n(q)$ versus q in the intermittency region. The A_1 value used is 1.0922 and $n = 10$. The $\sigma_n(q)$ has two peaks at $q = q_\alpha \approx 1.75$ and $q = q_\beta \approx 0.5$. Figure 2b shows the result for $A_1 = 1.096$ far after intermittency at which fully developed chaotic motion is observed. $\sigma_n(q)$ has only one peak at $q = q_\alpha \approx 0.6$.

A crisis, or sudden expansion in the size of the attractor is observed at $A = A_w \approx 0.747486$. For $A_1 = 0.73$, far from the crisis, $\sigma_n(q)$ has only one peak at $q = q_\alpha \approx 1.55$. Just before the crisis, corresponding to $A_1 = 0.747482$ the variance $\sigma_n(q)$ has two peaks at $q = q_\alpha \approx 2$ and $q_\beta \approx 0$.

From the above analysis we note that q_α peak is observed for all chaotic attractors. However, additional peaks in $\sigma_n(q)$ versus q plot occur just before and after the critical bifurcations. Thus it turns out that $\sigma_n(q)$ is useful for characterizing the chaotic attractors at bifurcations.

3. Characteristics of probability distribution in chaotic attractors of BVP oscillator

In this section we illustrate the characterization of weak and strong chaos in the BVP oscillator using the probability distribution of the k-step difference quantity of the variable V. Let $\{V_n\}$, $n = 1, 2, \ldots, N$ be the values of the variable V of (1) in the Poincaré map. We define the k-step difference quantity ΔV_k as $\Delta V_k = V_{m+k} - V_m$ where $m = 1, 2, \ldots, N', N' < N$. Then we calculate the probability distribution of ΔV_k.

Figure 3 shows the variation of maximal Lyapunov exponent (λ) as a function of A_1. In this plot there are many regions labelled by $\downarrow$, in which λ is positive but very small, indicating weak chaos. Distributions are calculated for various values of A_1. In numerical calculations we neglect the first 5000 Poincaré data as transient and use the next 10^4 points. For $A_1 = 1.0919$ the maximal Lyapunov exponent is ≈ 0.0004 and the corresponding chaotic motion is weak. For this A_1 value, P is found to change continuously with k. That is, the distribution is nonstationary. This is further verified by the chi-square, C^2 test quantity, defined as [15]

$$C^2(k,j) = \sum_{i=1}^{n} (R_i - S_i)^2/(R_i + S_i) \tag{7}$$

where R_i and S_i are the probabilities of the ith interval for $P(\Delta V_{k+j})$ and $P(\Delta V_k)$ respectively. In (7) intervals with $R_i = S_i = 0$ are excluded. If two probability distributions differ greatly, one gets a large C^2. For two similar distributions C^2 will be small. The numerically calculated C^2 is plotted in figure 4. The analysis is carried out for k values up to 5000. The nondecreasing C^2 implies the nonstationary characteristic of P. Thus, a complete probability distribution is impossible for weak chaos. A physical mechanism of nonstationary P can be a recurrence of memory loss and recovery of initial conditions [49].

Stationary probability distribution is found for strong chaos. For $A_1 = 1.26, \lambda$ is ≈ 0.056. The calculated C^2 is plotted in figure 5. For $k > 15$ C^2 is almost zero. That is,

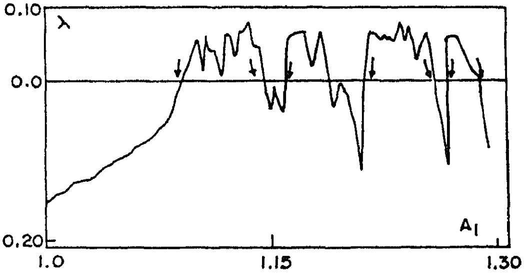


Figure 3. Estimated maximal Lyapunov exponent versus A_1.

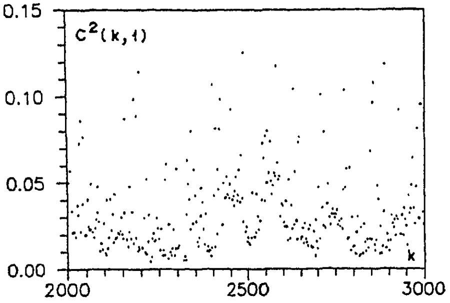


Figure 4. $C^2(k,1)$ versus k for the BVP chaotic attractor with $A_1 = 1.0919$ (weak chaos).

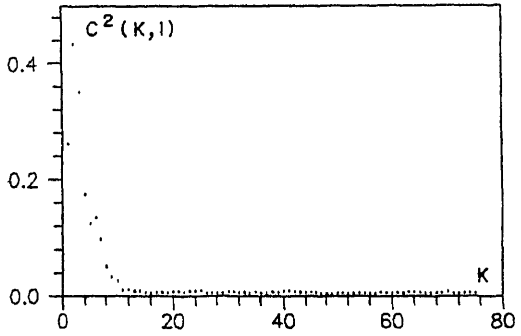


Figure 5. Same as figure (4) now for $A_1 = 1.26$ (strong chaos).

the distribution has evolved into a stationary state. This suggests that the variable V_n can be described as if it were generated by a random number generator with a certain probability distribution. A possible mechanism for a stationary distribution is complete loss of memory to the initial conditions. For two band and four band chaotic attractors, $C^2(k,1)$ values are found to be large (not shown). However, for two band and four band chaotic attractors $C^2(k,2)$ and $C^2(k,4)$ respectively decay to zero as k increases. Interestingly, for two (four) band attractor simple switching between two (four) different classes of P is found.

4. Controlling of chaos in the BVP equation

In this section we consider the problem of controlling of chaos in the BVP equation. First, we briefly review some previous results of controlling of chaos in the BVP equation.

4.1 *Previous results*

In a series of papers [27–33] we have studied the applicability and efficacy of a few control algorithms in the BVP equation.

Stabilization of UPO embedded in a chaotic attractor has been studied [28, 29, 31] by the control methods of Singer *et al* [21], Pyragas [22] and Chen and Dong [23]. A detailed comparative study of these methods has been made [31]. In all the three control schemes (i) stabilization of desired UPO is achieved for certain range of values of amplitude (ϵ) of the perturbation, (ii) before reaching the desired target the system exhibits transient evolution and (iii) the recovery time varies with the control strength ϵ.

Adaptive control algorithm [18, 19] with linear as well as nonlinear control functions is able to convert the chaotic attractors of the BVP system to a period-nT limit cycle [28, 29]. Stable control is observed for certain range of values of stiffness (ϵ) of the control function. Further, the recovery time is found to be inversely proportional to ϵ.

We have investigated analytically [30] and numerically [28] the effect of periodic parametric perturbation and addition of second periodic force on chaotic dynamics. The period of the controlled motion is found to depend on amplitude and frequency of the added force. We have shown that most of the regular regime can be identified by the Melnikov method. Interestingly, suppression of chaos is found in the parametric regimes where the Melnikov function does not change sign.

4.2 *Ott–Grebogi–Yorke method*

In the OGY technique [17], stabilization of an UPO in a suitably defined Poincaré map can be achieved by slight adjustment of a control parameter. Suppose we wish to stabilize an unstable period-1 orbit contained in the chaotic attractor of BVP oscillator. We choose A_0 as the control parameter and denote $X_F(A_0 = 0)$ as the unstable period-1 orbit to be stabilized. As A_0 in (1) is varied slightly from $A_0 = 0$ to some $A_0 \neq 0$ then $X_F(0)$ will shift to $X_F(A_0)$. We define a vector g as $g = \partial X_F(A_0)/\partial A_0$ evaluated at $A_0 \neq 0$. Near $X_F(0)$ and for small values of A_0 one may write $\delta X_{n+1} \approx M\delta X_n$ where M is a 2×2 matrix and $\delta X_n = X_n - X_F(A_0)$. When the nth iteration X_n is close to X_F the value of A_0 is changed such that X_{n+1} falls on the stable manifold of X_F. That is, we choose the control δA_{0n} so that $f_u \cdot \delta X_{n+1} = 0$ which yields the control formula [17]

$$\delta A_{0n} = \frac{\lambda_u f_u \cdot \delta X_n}{(\lambda_u - 1) f_u \cdot g}, \tag{8}$$

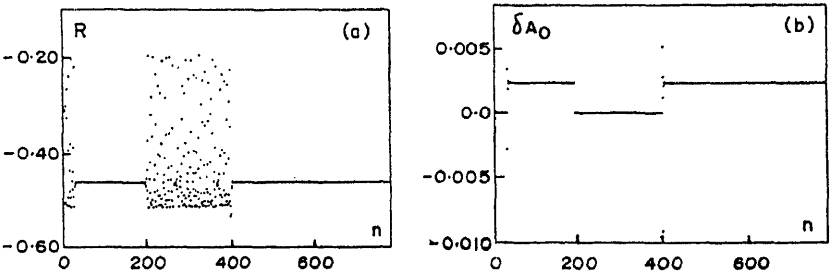


Figure 6. (a) The R-component of the period-1 orbit in the Poincaré map versus n. The control was switched on at $n = 30$, then switched off at $n = 200$ and then

where λ_u is the unstable eigenvalue of the matrix M and f_u is the unstable contravariant basis vector [17].

In the numerical simulation we found all pairs of iterates falling within the circles of radius 0.01 about the point $X_F(0)$. To this data we have fitted the approximate linear map and then determined the form of M and its eigenvalues and f_u. The approximate location of the period-1 unstable orbit is calculated as $X_F = (V_F, R_F) \approx (-0.9560, -0.4685)$. We have obtained $\lambda_u \approx 1.46, \lambda_s = -0.68, g \approx (0.105, 0.97)$ and $f_u = (1+\lambda_u^2)^{1/2}\,(\lambda_u - \lambda_s)^{-1}(1, \lambda_s)$. To stabilize the orbit X_F we have chosen the maximum allowed perturbation as $|\delta A_{0\,\max}| = 0.01$. Figure 6a illustrates the control of period-1 orbit. The period-1 orbit gets stabilized as long as the control is included. Figure 6b shows the perturbation δA_0 versus n.

4.3 *Suppression of chaos by periodic instantaneous burst*

The BVP equation with the addition of periodic instantaneous burst is written as [33]

$$\dot{V} = V - V^3/3 - R + A_1 \cos t, \tag{9a}$$

$$\dot{R} = c(V + a - bR) + \alpha g(t), \tag{9b}$$

where

$$g(t) = \sum_{n=1}^{\infty} \delta(t - n\tau T). \tag{9c}$$

The force $g(t)$ is nonzero and equal to unity only at times $t = n\tau T, n = 1, 2, \ldots$. For simplicity, T is fixed as 2π, period of the driving force. For fixed value of T the times at which $g(t)$ become nonzero and the period of the force are characterized by τ. Equation (9) is integrated using fourth order Runge–Kutta method with time step $2\pi/100$.

Figure 7 shows the bifurcation diagram as a function of α for $\tau = 0.5$. Chaotic behaviour persists for $\alpha \leq 0.0045$. Regular motion is recovered in the parameter interval $\alpha > 0.0045$. An interesting point here is the suppression of chaos by inverse period doubling phenomenon. Period $8T$, $4T$, $2T$ and T motions are found in the intervals (0.0051–0.0057), (0.006–0.0096), (0.0102–0.0309) and ($\alpha > 0.0312$) respectively. Next,

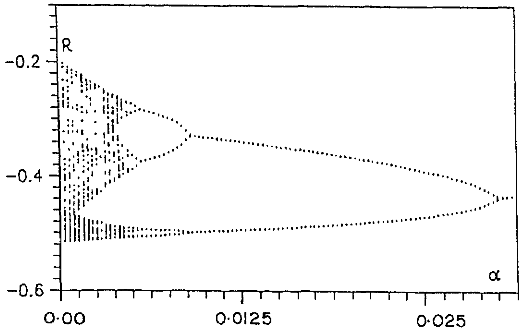


Figure 7. Bifurcation diagram of the BVP system (9). The parameters are fixed at $a = 0.7$, $b = 0.8$, $c = 0.1$, $A_1 = 0.74$, $T = 2\pi$ and $\tau = 0.5$.

α is kept at 0.006286 and τ is increased from small value in steps of 0.01 up to 1. Regular behavior is observed for $0.01 \leq \tau \leq 0.69$.

Further, the effect of phase difference ϕ between the periodic external force $A_1 \cos t$ and $g(t)$ is also studied. For various values of τ and α the dynamics of the BVP system is studied by varying ϕ. Interestingly, the behaviour of the system whether it is chaotic or periodic in the absence of ϕ is unaltered when ϕ is included. The only effect observed is an infinitesimal shift in the values of the state variables in the Poincaré map. A more detailed investigation of suppression of chaos by instantaneous burst will be presented elsewhere [33].

5. Summary and conclusions

In this paper we have studied various aspects of chaotic dynamics in the BVP oscillator. As shown for the BVP oscillator the chaotic attractors just near the critical bifurcations such as sudden widening, band merging and intermittency produce large fluctuations of the coarse-grained local expansion rates $\wedge$, and consequently, additional peaks are found to occur in the $\sigma_n(q)$ versus q plot. On the other hand, a stationary probability distribution is found for chaotic attractors with large positive Lyapunov exponent, while for attractors with sufficiently small positive Lyapunov exponent a nonstationary probability distribution is observed. Thus, $\sigma_n(q)$ and $P(\Delta x_k)$ are useful to characterize various chaotic attractors.

We have discussed the problem of controlling of chaos in the BVP equation. If the aim is to set the system motion to a desired predetermined regular orbit, the adaptive control algorithm and feedback methods have distinct advantages. If one wishes to convert chaotic motion to some regular motion then periodic perturbations can be used. Surprisingly, chaotic motion can be suppressed by periodic kicks, which make the motion periodic.

References

[1] H G Schüster, *Deterministic chaos* (Verlag Chemie, Weinheim, 1988)
[2] M Lakshmanan and K Murali, *Chaos in nonlinear oscillators: controlling and synchronization* (World Scientific, Singapore, 1996)
[3] H D I Abarbanel, R Brown and M B Kennel, *J. Nonlinear. Sci.* **1**, 175 (1991)
[4] C Amitrano and R S Berry, *Phys. Rev.* **E47**, 3158 (1993)
[5] N Voglis and G J Contopoulos, *J. Phys.* **A27**, 4899 (1994)
[6] J L Chern and K Otsuka, *Phys. Lett.* **A188**, 321 (1994)
[7] S K Nayak, R Ramaswamy and C Chakravarty, *Phys. Rev.* **E51**, 3376 (1995)
[8] S Rajasekar, *Indian J. Phys.* **B70**, 51 (1996)
[9] R Ishizaki, T Horita and H Mori, *Prog. Theor. Phys.* **89**, 947 (1993)
[10] H Hate, T Horita and H Mori, *Prog. Theor. Phys.* **82**, 897 (1989)
[11] T Horita and H Mori, *Prog. Theor. Phys.* **91**, 677 (1994)
[12] S Rajasekar, *Phys. Rev.* **E52**, 3234 (1995)
[13] S Rajasekar, *Chaos, Solitons and Fractals* (1996, in press)
[14] P Philominathan and S Rajasekar, *Physica* **229**, 244 (1996)
[15] J L Chern and T C Chow, *Phys. Lett.* **A192**, 34 (1994)
[16] S Rajasekar, *Characteristics of probability distribution in chaos in an one dimensional map and BVP oscillator* (in preparation)

[17] E Ott, C Grebogi and J Yorke, *Phys. Rev. Lett.* **64**, 1196 (1990)
[18] B A Huberman and E Lumer, *IEEE Trans. Circuits Syst.* **37**, 547 (1990)
[19] S Sinha, R Ramaswamy and J Subba Rao, *Physica* **D43**, 118 (1990)
[20] R Lima and M Pettini, *Phys. Rev.* **A41**, 726 (1990)
[21] J Singer, Y Z Wang and H H Bau, *Phys. Rev. Lett.* **66**, 1123 (1991)
[22] K Pyragas, *Phys. Lett.* **A170**, 421 (1992)
[23] G Chen and X Dong, *Int. J. Bifur. Chaos* **2**, 407 (1992)
[24] Y Braiman and I Goldhirsch, *Phys. Rev. Lett.* **66**, 2545 (1991)
[25] A Hübler and E Luscher, *Naturwissenschaften* **76**, 67 (1989)
[26] E A Jackson, *Phys. Rev.* **A44**, 4839 (1991)
[27] S Rajasekar and M Lakshmanan, *Int. J. Bifur. Chaos* **2**, 201 (1992)
[28] S Rajasekar and M Lakshmanan, *Physica* **D67**, 282 (1993)
[29] S Rajasekar and M Lakshmanan, *J. Theor. Biol.* **166**, 275 (1994)
[30] S Rajasekar, *Pramana – J. Phys.*, **41**, 295 (1993)
[31] S Rajasekar, *Chaos, solitons and fractals*, **5**, 2135 (1995)
[32] S Rajasekar, *Phys. Rev.* **E51**, 775 (1995)
[33] S Rajasekar, *Controlling of chaos by periodic instantaneous burst* (submitted for publication)
[34] J K John and R E Amritkar, *Int. J. Bifur. Chaos* **4**, 1687 (1994)
[35] J K John and R E Amritkar, *Phys. Rev.* **E49**, 4843 (1994)
[36] K Murali and M Lakshmanan, *J. Circuits Systems and Computers* **3**, 125 (1993)
[37] S Parthasarathy and S Sinha, *Phys. Rev.* **E51**, 6239 (1995)
[38] S D Gadre and V S Varma, *Pramana – J. Phys.* **45**, 355 (1995)
[39] G Chen and X Dong, *Int. J. Bifur. Chaos* **3**, 1363 (1993)
[40] L M Pecora and T L Carroll, *Phys. Rev. Lett.* **64**, 821 (1990)
[41] R E Amritkar and Neelima Gupte, *Phys. Rev.* **A44**, 3403 (1991)
[42] Neelima Gupte and R E Amritkar, *Phys. Rev.* **E48**, R1620 (1993)
[43] R E Amritkar and Neelima Gupte, *Phys. Rev.* **E47**, 3889 (1993)
[44] K Murali and M Lakshmanan, *Phys. Rev.* **E48**, R1624 (1993)
[45] K Murali, M Lakshmanan and L O Chua, *Int. J. Bifur. Chaos* **5**, 563 (1995)
[46] A C Scott, *Neurophysics* (Wiley, New York, 1977)
[47] S Rajasekar and M Lakshmanan, *Physica* **D32**, 146 (1988)
[48] S Rajasekar, S Parthasarathy and M Lakshmanan, *Chaos, Solitons and Fractals* **2**, 271 (1992)
[49] P M Gade and R E Amritkar, *Phys. Rev.* **A45**, 725 (1992)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 1,
January 1997
pp. 259–270

Control of chaos

SANGEETA D GADRE and V S VARMA
Department of Physics and Astrophysics, University of Delhi, Delhi 110 007, India

Abstract. We review the subject of control of chaotic systems paying special attention to exponential control. We also discuss the application of synchronization of chaotic systems to security in communications.

Keywords. Exponential control; synchronization; secure communications.

PACS No. 05.45

1. Introduction

This started off as a general review of the subject of synchronization and control of chaotic systems. When it became clear however that this special issue on chaos would contain a number of such contributions, we decided to give it a more personal touch by also discussing at some length some of our own work in this field – the exponential control of systems exhibiting chaotic behaviour.

2. Control

The subject of controlling chaotic systems has become popular since the work of Ott *et al* [1], hereafter referred to as OGY. OGY were the first to realize that the presence of chaos instead of being a nuisance can, in fact, be of great advantage. This is because a chaotic attractor typically has embedded within it an infinite number of unstable periodic orbits. Any one of a number of such orbits can be stabilized by small time-dependent parameter perturbations of an available system parameter and a choice can be made among these orbits to achieve a desired system performance. This is to be contrasted with the situation in which the attractor is not chaotic but periodic, in which case small parameter perturbations can only produce small changes in the system orbit.

To appreciate the implications of what has been said above, consider a system which may be required to be used for different purposes and be expected to operate under different conditions at different times. If the system is chaotic, this type of multiple-use situation might be accommodated without altering the gross system configuration. In particular, depending on the use desired, the system behaviour can be changed by switching the temporal programming of the small parameter perturbations so as to purposely building chaotic dynamics into the system can be a way of accommodating the desired flexibility.

OGY imagined a situation in which the dynamical equations describing the system are not known, but the experimental time-series of some scalar variable can be measured and using delay coordinates [2] a surface of section can be constructed. The piercings of the experimental surface of section are then studied and a particular unstable periodic orbit which gives the best system performance is selected (using embedding techniques discussed by Gunaratne *et al* [3]). The system trajectory is observed till it falls within a strip of predefined width of the selected unstable orbit, at which juncture a suitable parameter p of the system is adjusted so that the system trajectory at the next piercing of the surface of section falls on the stable manifold of the selected unstable orbit. When this occurs, the parameter perturbation can be set to zero and the orbit for subsequent time will approach the desired fixed point at a geometrical rate.

OGY demonstrated the success of their analysis numerically on the Hénon map. The method has subsequently been successfully applied to a variety of systems including magnetic ribbons [4], spin-wave systems [5], chemical systems [6], electrical diodes [7], lasers [8] and cardiac systems [9].

Some important features of the OGY technique are:

1. no model for the dynamics of the system under consideration is required,
2. the required changes in the parameter can be quite small,
3. control can be achieved even with imprecise measurements of the system eigenvalues and eigenvectors, and
4. the method is applicable to any system whose dynamics can be characterized by a nonlinear map.

Shinbrot *et al* [10, 11] adapted the OGY technique to situations in which the system can be modelled by a set of equations although these need not be exact. Their technique has been shown to be effective even in the presence of small-amplitude noise or small errors in modelling.

Dressler and Nitsche [13] showed that modifications to the original OGY procedure were necessary when the attractor is reconstructed from a time series using delay coordinates because the experimental surface of section map depends not only on the current value of the control parameter but also on its preceding value. Rollins *et al* [14] presented a recursive proportional-feedback algorithm, which is an adaptation of the method of Dressler and Nitsche for the control of highly dissipative systems.

Lai *et al* [15] extended the OGY stabilization of dissipative chaos to chaotic Hamiltonian systems. Lai and Grebogi [16] used this modification to synchronize two signals using perturbations proportional to the instantaneous difference between them. Alvarez-Ramírez [17] presented a nonlinear feedback control in which the feedback is based on linearizing the input-output dynamics of the system and control leads to large regions of stability.

Newell *et al* [18] demonstrated an experimental proportional-feedback algorithm for the synchronization of chaotic time signals generated by a pair of independent diode resonator circuits.

Bielawski *et al* [19] stabilized unstable periodic orbits by applying continuous feedback on a control parameter, the feedback signal being proportional to the difference between two values of a dynamical variable separated by a time interval equal to the unstable orbit periodicity. This method is expected to be efficient for systems near a bifurcation point and for highly dissipative systems.

Pyragas [20] proposed two methods for the control of unstable periodic orbits with small time-continuous perturbations. Feedback combined with a periodic external force of a special form is used in the first method (external force control); while the second method relies on self controlling delayed feedback (delayed feedback control). Qu *et al* [21] analyzed Pyragas' external force control and found that continuous control by feedback of a single variable is not always successful. They also showed, contrary to Pyragas' claim, that under certain conditions positive feedback can control a system where negative feedback fails to do so. Pyragas [22] subsequently extended his control to stabilize aperiodic orbits, while Kittle *et al* [23] verified the effectiveness of such control experimentally.

Matias and Güémez [24, 25] presented a new method for the stabilization of 1-D iterated maps which exhibit chaos, by acting on the system variables rather than on some available parameter. The method is implemented by applying pulses that change the system variable proportionally. The control algorithm consists of the application every Δn iterations of a feedback to the variable x having the form

$$x_n = x_n(1+\gamma),$$

where γ represents the strength of the feedback and can be positive as well as negative. The method does not require that the particular dynamical law be known. However, the ability to control chaos depends critically on the values of γ and Δn. The latter is closely related to the periodicity of the observed cycle, this being a multiple of Δn. They illustrated their technique by an application to the logistic and exponential maps. Subsequently they applied their method to the isothermal autocatalator model [26] and the Lorenz and Rössler systems.

Singer *et al* [27] demonstrated experimentally and theoretically that low-energy, feedback control signals can be utilized efficiently to suppress (laminarize) chaotic flow in a thermal convection loop. Chen and Chou [28] showed that the continuous feedback approach is highly effective for controlling chaotic systems. Given the system's model dynamic equations, the control is designed for a well-specified system performance. The approach was applied to the Lorenz system and was able to drive the system to any steady state. The control was found to tolerate both measurement noise and modeling uncertainty as long as these were bounded.

Huberman and Lumer [12] introduced a simple adaptive control mechanism to control nonlinear systems in stable orbits. A system which is perturbed away from a stable fixed point (say X_s^*) is brought back to the fixed point by introducing changes in the parameter (μ) such that

$$\mu_{n+1} = \mu_n + \epsilon(X_n - X_s^*).$$

Sinha *et al* [29] extended this method to multi-parameter and higher dimensional nonlinear systems. They also suggested a method by which periodic motion such as limit cycles can be adaptively controlled and demonstrated the robustness of the procedure in the presence of (additive) background noise.

John and Amritkar [30] suggested a method for controlling chaotic systems which is able to synchronize the system trajectory to a desired unstable orbit. Their method uses adaptive control [12] and introduces time-dependent changes in the system parameters depending on two factors : (1) the difference between the system output variables $\{u_j\}$ and the corresponding variables of the desired orbit $\{v_j\}$, and (2) the difference between the values of the parameters (μ_i) which are controlled and their values (μ_i^*) for the desired orbit. They demonstrated the effectiveness of their method using the Lorenz and Rössler systems. They also showed that their method could be used for communication purposes.

There are other interesting and simple methods [31–38] for controlling of chaos. For example, entrainment control [31], suppression of chaos by weak periodic parametric perturbation [32] and addition of second periodic force [33] are also possible.

3. Synchronization

Although we have mentioned synchronization en passant, the possibility of synchronization in chaotic systems is not obvious. Two identical autonomous chaotic systems started at nearly identical initial points in phase space have trajectories which quickly become uncorrelated, even though each maps out the same attractor in phase space. Pecora and Carroll [39] were the first to describe the linking of two chaotic systems with a common signal such that the two systems synchronized, i.e. the trajectories of the two systems converged and remained in step. Their results were obtained for flows (differential equations), but can be applied to iterated maps with only small modifications. They considered an autonomous n-dimensional dynamical system

$$\dot{u} = f(u),$$

divided the system, arbitrarily, into two subsystems $[u = (v, w)]$. They created a new system u' identical to the original u system and substituted the set of variables v for the corresponding v' in the new system. The behaviour of the new system depends on the Lyapunov exponents of the w subsystem—referred to as the sub-Lyapunov exponents (SLEs). They suggested the following theorem: the subsystems w and w' will synchronize only if the SLEs are all negative.

This theorem is necessary, but not sufficient, for synchronization. It says nothing about the set of "initial conditions" for which w' will synchronize with w. They demonstrated synchronization in the Lorenz and the Rössler systems as well as in electronic circuits and their numerical models [40]. Incidentally, Amritkar and Gupte [41] have proposed a qualitative measure of correlation between coevolving dynamical systems which is very useful in characterizing synchronization in chaotic systems.

Periodically driven nonlinear systems can exhibit multiple-period behaviour (period-2, period-3, etc.). Several such systems, when driven by the same drive, can be on identical attractors but remain out of phase with each other. This means that the basins of attraction for multiple-period systems can be divided into domains of attraction, one for each phase of the motion. A period-n attractor will have n domains of attraction in its basin. This out-of-phase situation is stable—small perturbations will not succeed in getting the systems in phase. Pecora and Carroll [42] showed that by using an almost periodic driving signal (generated from various chaotic systems) one can simultaneously (1) keep the motion of the systems nearly the same as the periodic driving signal, (2) keep the basins of attraction nearly the same, and (3) eliminate the n domains of attraction. In other words, there will be only one domain for the basin. This means that any number of such driven systems will always be in phase.

Kittle *et al* [23] pointed out that the stabilization of a desired aperiodic orbit can be interpreted in terms of Pecora and Carroll's method of synchronization. To stabilize a desired orbit we need to know a prerecorded signal $y^*(t)$, which is sufficient to control the output of a given system to a desired orbit for a given controlling method. The prerecorded signal $y^*(t)$ can be replaced by the output signal of an additional, identical chaotic system. The latter can be seen as the driving system, and the original can be interpreted as the response system. The problem of stabilizing a prerecorded desired aperiodic orbit is thus equivalent to the problem of synchronizing two identical chaotic systems.

Kapitaniak [43] applied the continuous controlling method developed by Pyragas [20] to achieve synchronization of two chaotic systems. Murali and Lakshmanan [44] also used Pyragas' method to synchronize two identical nonlinear oscillators and to transmit signals in a secure way.

Ding and Ott [45] pointed out that exact synchronism may also occur for a large class of systems that are not replicas of the original system as suggested by Pecora and Carroll. They were able to affect the following improvements over Pecora and Carroll's technique:

1. achieving synchronization when a replica subsystem does not synchronize (i.e. $\Lambda > 0$ for the replica subsystem),
2. enabling faster convergence to the synchronized state,
3. eliminating or reducing the size of spurious subsystem basins of attraction in which the subsystem does not synchronize, and
4. improving the performance of signal recovery techniques for situations where a chaotic time series is used to mask a small information bearing signal [44, 46].

Gupte and Amritkar [47] demonstrated that it is possible to stabilize unstable periodic orbits of chaotic attractors by using a suitable drive variable, the choice of drive variable being dictated by the values of the SLEs of the response system. They found that for the Lorenz and Rössler systems, the SLEs of some of the unstable fixed points appeared to govern the locking to chaotic orbits and to periodic orbits as well. On the other hand, the SLEs of the unstable period-six orbit appeared to govern the properties of the chaotic orbit for the Duffing oscillator. They also noted that the actual chaotic trajectories for these systems wound around the unstable fixed points or the periodic orbits for which the agreement between the SLEs was observed. They conjectured that the SLEs of a given chaotic orbit retained the memory of the unstable periods that were responsible for its origin.

4. Exponential control

We [48, 49] have succeeded in implementing a somewhat novel method of controlling chaos. We impose a multiplicative functional feedback control on a system parameter.

The control is of the exponential form whose argument is proportional to the feedback response of the system, i.e. the difference between the desired value and the actual value of one suitably chosen variable of the system. Pecora and Carroll's method uses a suitably chosen variable of the system as a drive for controlling chaos. The OGY method controls chaos by perturbing the system parameter. Our method is a combination of variable and parametric control as it achieves control by modulating the parameter through the variable. The control can stabilize unstable fixed points, unstable limit cycles as well as chaotic trajectories. It is effective both for maps and flows.

Consider a general N-dimensional dynamical system

$$\dot{\vec{X}} = \vec{F}(\vec{X}; \mu; t) \tag{1}$$

where $\vec{X} \equiv (X_1, X_2, \ldots.., X_N)$ are variables and $\mu \equiv (\mu_1, \mu_2, \ldots., \mu_K)$ are controlling parameters whose values determine the nature of the dynamics. The stabilization of a desired unstable attractor or a chaotic trajectory is possible by multiplying a suitably chosen controlling parameter say μ_r in (1) by an exponential feedback control involving only one suitably chosen variable say X_l such that the form of control is

$$\exp[\epsilon(X_l - X_l^s)] \tag{2}$$

where X_l is the actual value of one of the variables of the system after applying the control, X_l^s is the desired value of that variable and ϵ is the "stiffness" of the control which can take both positive and negative values.

The dynamics of the modulated system in the presence of control is given by

$$\dot{\vec{X}} = \vec{F}(\vec{X}; \mu_1, \mu_2, \ldots, \mu_r \exp[\epsilon(X_l - X_l^s)], \ldots, \mu_K; t). \tag{3}$$

The form of the control is such that it becomes passive once the desired goal $\overrightarrow{X^s}$ is achieved. If fluctuations drive the system off the desired goal, the control reactivates.

The control works for those combinations of controlling parameters and variables of the system for which the real part of the largest Lyapunov exponent of the modulated system, the system after the control is applied represented by (3), is negative. The feedback function in the expression of the control involves only one selected variable X_l to convert the desired repellor – a fixed point, a limit cycle or a chaotic orbit, into an attractor, implying that knowledge of only one variable on the desired unstable orbit is sufficient to settle the system on to that orbit. This makes exponential control of particular significance for systems with several degrees of freedom. As an example, one unstable fixed point of the Lorenz system [50] can be stabilized using exponential control with feedback depending on any one of the X, Y or Z variables, whereas an unstable limit cycle or a chaotic trajectory of the system can be converted into an attractor only if we use feedback depending on the Z variable. For stabilizing most repellors, it is sufficient to apply exponential control to one selected parameter of the Lorenz system. But for stabilizing one of the two fixed points of the Hénon map [51], exponential control has to be applied to both the parameters a and b. Some of these features may also be true of other forms of control. Although in all the cases we have studied, control involving only one variable on the desired orbit is sufficient to make all the Lyapunov exponents of the system negative, this may not be true for all systems. But the fact remains that the control uses only a subset of the variables and the parameters for controlling chaos. We have successfully tested our control for stabilizing different types of orbits of the logistic map, the Hénon map and the Lorenz system.

We can extend our control to convert unstable fixed points of higher period say $(\overrightarrow{X^{*1}}, \overrightarrow{X^{*2}}, \ldots, \overrightarrow{X^{*k}})$ to stable fixed points for given values of the controlling parameters. What is required is a feedback that encodes as much information about a periodic orbit as is necessary for its unique characterization. But there is a problem here of practicality. Since in such cases the form of the control given by (2) requires the convergence of the chosen variable to one of the k values of that variable, the controlling technique diminishes in utility with increase in period. For higher period orbits of a discrete dynamical system a more effective control is one which employs a logical OR structure in the feedback function. So in order to stabilize unstable fixed points of period k, i.e. $\{\overrightarrow{X^{*i}}\}$, one of the parameters μ_r in (1) is multiplied by

$$\exp\left(\epsilon \prod_{i=1}^{k} (X_l(n) - X_l^{*i})\right) \tag{4}$$

where $X_l(n)$ is the value of the chosen variable X_l at time n of the modulated map after applying control and $(X_l{}^{*1}, X_l{}^{*2}, \ldots., X_l{}^{*k})$ is the set of k values of X_l on the desired unstable period-k orbit. We have successfully implemented such control in the case of the logistic map for stabilizing a period-2 orbit [48].

A quantity of obvious interest in the context of controlling chaos is the time required for the system to settle on to the desired orbit. This of course depends upon the stiffness of the control i.e. ϵ. For a given ϵ we study the time τ required for the system to approach within a distance ω of the desired orbit starting from some initial point. If ω_0 is the initial distance from the desired orbit, then it is clear that the length of the transient τ and ω are related by $\omega = \omega_0 \exp(\lambda\tau)$, where λ is the real part of the largest Lyapunov exponent. The slope of the plot of τ against $\ln(\omega/\omega_0)$ is nothing but $1/\lambda$. This of course has to be negative for convergence. These values of the Lyapunov exponents are found to be in good agreement with those obtained numerically [52] and, wherever possible, analytically. We also study the dependence on ϵ of the transient time τ required for settling on to different types of orbits for different systems characterized by $\ln(\omega/\omega_0)$ as well as on the accuracy ω with which it does so. We find that τ is initially a decreasing function of ϵ. But there exists an optimum stiffness of control beyond which increasing ϵ can cause τ to increase. This dependence of τ on ϵ is found to be the same as that of λ on ϵ for any orbit.

In addition to this, in the case of discrete maps we have also tried to stabilize orbits which are not the natural fixed points (stable or unstable) of the original system. We have found that exponential control succeeds in creating new stable attractors which are not the natural attractors of the unmodulated system. However we cannot stabilize any arbitrary orbit. The functional form of the map and the control criterion decide which orbits can be forced on to the system.

There is one drawback of the control – in any system, if an unstable fixed point is represented by a null vector then it cannot be stabilized using exponential control. The reason is easy to understand. The Jacobian of the modulated system given by (3) evaluated at such a point in the presence of the control is the same as of the unmodulated system (1). So the eigenvalues of the system remain unchanged in the presence of the

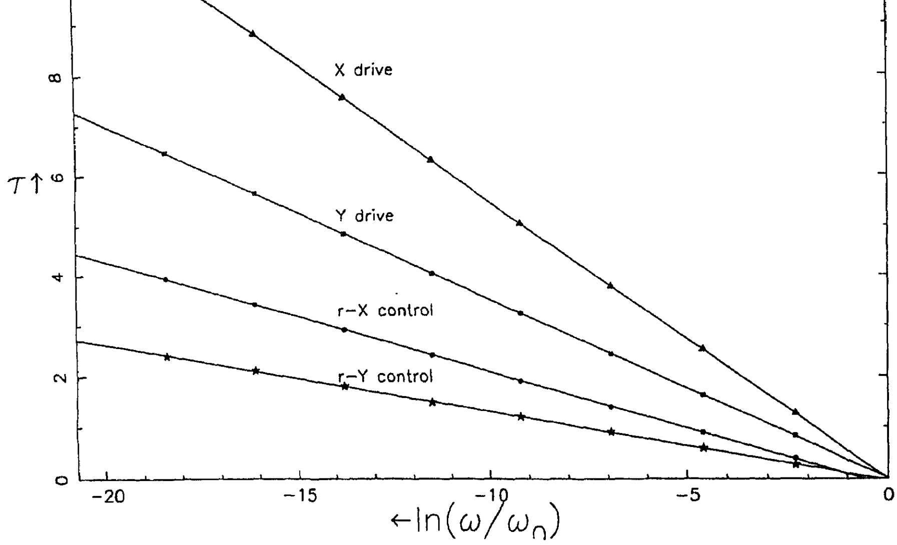


Figure 1. Plot of the transient time τ vs $\ln(\omega/\omega_0)$ for the Lorenz attractor with $\sigma = 10$, $b = 8/3$ and $r = 60$ while stabilizing the unstable fixed point $(\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1)$. Values of τ for Pecora and Carroll's method with X-drive (triangles) and Y-drive (squares) are compared with those obtained by using exponential control involving the parameter r and the variable X (circles) and also r and y (stars).

control. Hence such a fixed point remains unstable even in the presence of control. Such points may, however, get stabilized in the process of stabilizing other fixed points, resulting in the coexistence of more than one attractor.

We have compared the transient times and the basins of attraction for exponential control with other forms of control which share with it the following features:

1. the control is continuous,
2. it depends only on the variable(s) of the desired orbit,
3. is independent of the previous values of the controlling parameters μ, if the values of the parameters are changing in the presence of control, and
4. does not increase the dimensionality of the original system.

To compare our control with Pecora and Carroll's, both for X- and Y-drives in the Lorenz system, we consider the control where the parameter r is multiplied by the exponential feedback function involving the X and Y variables keeping the same set of parameter values in both cases. We find that the value of Λ with this form of the control is $-4.536\cdots$ for $\epsilon = 1.7$. We also find that Λ for the control involving the parameter r and the variable Y is $-7.593\cdots$ for $\epsilon = -0.05$ for the same set of parameter values. The corresponding values of Λ for Pecora and Carroll's method are -1.83 for the X-drive and

-2.85 for the Y-drive [39]. Our values of Λ are less than both these values and hence exponential control can be considered to take hold more rapidly. To verify this we find the transient time τ, required to stabilize the second unstable fixed point of the Lorenz system for a given set of parameter values, for different values of ω using Pecora and Carroll's method and for exponential control involving the parameter r and the variables X and Y for $\epsilon = -1.7$ and $\epsilon = -0.05$ respectively. We plot the variation of τ with $\ln(\omega/\omega_0)$ in figure 1, which confirms that our control stabilizes the unstable fixed point faster than Pecora and Carroll's. This is not altogether surprising in view of the fact that our control has an exponential form.

We have also compared our control with those suggested by Singer *et al* [27], and Chen and Chou [28]. For the first method, we find that the basin of attraction, while stabilizing nonzero fixed points of the Lorenz system, is smaller than the one obtained using exponential control. Also, the transient time is of the order $\sim 10^5$ which is larger than that with exponential control ($\sim 10^3$). The method of Chen *et al* [28] shows the same transient time for stabilizing the unstable fixed points of the Lorenz map but the basin of attraction is smaller than that obtained with exponential control. These comparisons should not be taken to imply that exponential control will always be the better choice. They only show that the performance of exponential control is different from other controls and may be more efficient for some systems.

With exponential control it is also possible to control spatiotemporal chaos. A spatiotemporal system can be described by a coupled map lattice (CML) model [53–57]. Feedback pinnings [58] are used to control chaos in the system by stabilizing a certain unstable reference state. To control such a system F, we use pinnings defined by

$$X_{n+1}(i) = F \sum_{k=1}^{L/I} \delta(i - Ik - 1) \exp[\epsilon(X_n(i) - \overline{X}_n(i)] \tag{5}$$

where I is the distance between two neighbouring pinnings, $\overline{X}_n(i)$ is the reference state to be stabilized and $\delta(j) = 1$ for $j = 1$ and $\delta(j) = 0$ otherwise. We have succeeded in stabilizing various unstable spatiotemporal patterns like an unstable homogeneous stationary state, an unstable space-period-two pattern and a time-period-two with a space-period-four running wave.

5. Application of synchronized chaos to communication

Cuomo and Oppenheim [46] used an analog circuit implementation of the chaotic Lorenz system to illustrate the use of synchronized chaotic systems in communications. The set of equations representing Lorenz based chaotic circuits are

$$\begin{aligned}
\dot{u} &= \sigma(v - u) \\
\dot{v} &= ru - v - 20uw \\
\dot{w} &= 5uv - bw.
\end{aligned} \tag{6}$$

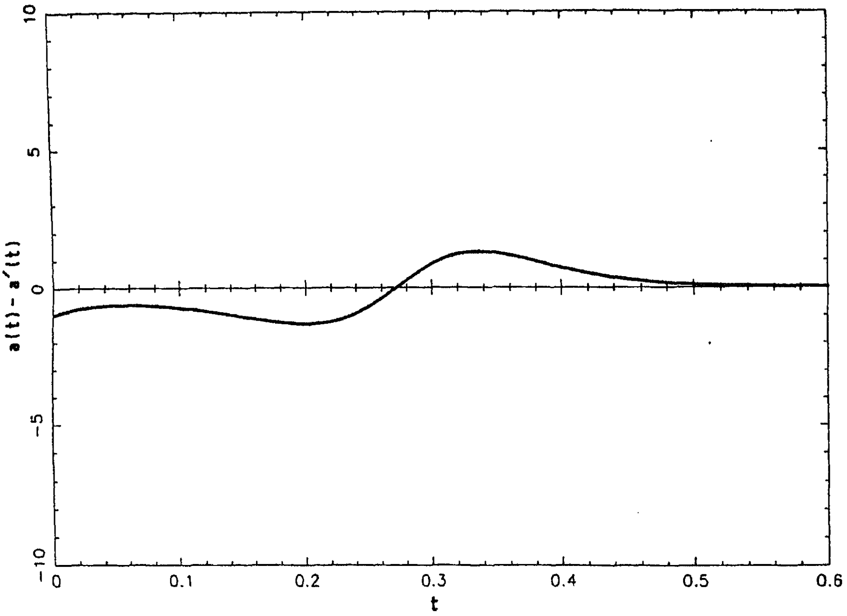


Figure 2. The error i.e. $a'(t) - a(t)$, in regeneration of the information signal at the receiver.

drive signal. The basic idea was to modulate a transmitter coefficient b with the information-bearing wave form, $m(t)$, which was taken to be a square wave. The square wave changed the value of b to a new value b_n with the one bit and no change was produced with the zero bit. A chaotic drive signal $u(t)$ was transmitted. At the receiver, the modulation produced a synchronization error between the received drive signal and the receiver's regenerated drive signal u_r with an error signal amplitude that depended on the modulation.

We have tried to use exponential control for the modulation/detection process in communications. The idea is to take the transmitter and receiver systems to be represented by the same set of equations with the same set of values for the controlling parameters. As demonstration we choose the Lorenz system to represent the transmitter and the receiver, with the parameter values $\sigma = 16.0$, $b = 4.0$, and $r = 40.0$. The largest Lyapunov exponent, for these values of the parameters is positive, implying that the system is in a chaotic regime. We chose an arbitrary chaotic trajectory starting with the initial point, say, $(10.0, 0.0, 30.0)$. The information signal $a(t)$ is taken to be a square wave defined by

$$\begin{aligned} a(t) &= 10 \qquad 0 \le t < 5\Delta t \\ &= 0 \qquad 5\Delta t \le t < 10\Delta t \end{aligned} \tag{7}$$

where Δt is the integration step taken to be 0.001. We superimpose the information signal $a(t)$ on one of the variables, say, $X(t)$ of the drive system such that the resultant signal $s(t) = a(t) + X(t)$. Then $s(t)$ and another variable, say, $Z(t)$ of the drive system are transmitted. At the receiver, we start with an arbitrary initial state, say, (X_r^0, Y_r^0, Z_r^0). We apply $b - Z$ control such that the parameter b gets multiplied by $\exp(\epsilon(Z_r - Z))$, where Z_r is the Z coordinate of a point on a chaotic trajectory of the receiver, evolving in the presence of exponential control, so that its dynamics can be represented by

$$\begin{aligned}\dot{X}_r &= \sigma(Y_r - X_r)\\ \dot{Y}_r &= -X_r Z_r + rX_r - Y_r\\ \dot{Z}_r &= X_r Y_r - b\exp(\epsilon(Z_r - Z))Z_r.\end{aligned} \tag{8}$$

We choose the value of ϵ to be 0.034 for which the transmitted chaotic trajectory becomes stable in the presence of exponential control. Consequently nearby trajectories converge to the desired trajectory. Thus after transients have died out, the information signal can be regenerated by subtracting the $X_r(t) \approx X(t)$ signal from the transmitted signal $s(t)$, i.e, $a(t) \approx a'(t) = s(t) - X_r(t)$. We find that the error in the regeneration of the information signal $a(t)$ for the chosen values of the constants, lies between 0.1% and 1.0% of the amplitude of the information signal. This can be seen in figure 2.

The advantage of using this approach for the modulation/detection of a signal is that there is no restriction on the frequency or the amplitude of the information signal. The quality of the regenerated signal is quite good. The only drawback is that we have and transmit two signals, viz. $s(t)$ and $Z(t)$, from the transmitter.

Acknowledgement

We would like to thank Ram Ramaswamy for inviting us to write this review.

References

[1] E Ott, C Grebogi and J A Yorke, *Phys. Rev. Lett.* **64**, 1196 (1990)

[2] F Takens, in *Dynamical systems and turbulence*, edited by D Rand and L S Young (Springer-Verlag, Berlin, 1981)

[3] G H Gunaratne, P S Linsay and M J Vinson, *Phys. Rev. Lett.* **63**, 1 (1989)

[4] W L Ditto, S N Rauseo and M L Spano, *Phys. Rev. Lett.* **65**, 3211 (1990)

[5] A Azevedo and S M Rezende, *Phys. Rev. Lett.* **66**, 1342 (1991)

[6] B Peng, V Petrov and K Showalter, *J. Phys. Chem.* **95**, 4957 (1991)

[7] E R Hunt, *Phys. Rev. Lett.* **67**, 1953 (1991)

[8] R Roy, T W Murphy Jr, T D Maier, Z Gills and E R Hunt, *Phys. Rev. Lett.* **68**, 1259 (1992)
C Reyl, L Flepp, R Badii and E Brun, *Phys. Rev.* **E47**, 267 (1993)

[9] A Garfinkel, W L Ditto, M L Spano and J Weiss, *Science* **257**, 1230 (1992)

[10] T Shinbrot, E Ott, C Grebogi and J A Yorke, *Phys. Rev. Lett.* **65**, 3215 (1990); *Phys. Rev.* **A45**, 4165 (1992)

[11] T Shinbrot, W Ditto, C Grebogi, E Ott, M Spano and J A Yorke, *Phys. Rev. Lett.* **68**, 2863 (1992)

[12] B A Huberman and E Lumer, *IEEE Trans. Circuit Syst.* **37**, 547 (1990)

[13] U Dressler and G Nitsche, *Phys. Rev. Lett.* **68**, 1 (1992)

[14] R W Rollins, P Parmananda and P Sherad, *Phys. Rev.* **E47**, R780 (1993)

[15] Y C Lai, M Ding and C Grebogi, *Phys. Rev.* **E47**, 86 (1993)

[16] Y C Lai, M Ding and C Grebogi, *Phys. Rev.* **E49**, 1094 (1994)

[17] J Alvarez-Ramírez, *Phys. Rev.* **E48**, 3165 (1994)

[18] T C Newell, P N Alsing, A Gavrielides and V Koranis, *Phys. Rev.* **E49**, 313 (1994)
[19] S Bielawski, D Derozier and P Gloreieux, *Phys. Rev.* **E49**, R971 (1994)
[20] K Pyragas, *Phys. Lett.* **A170**, 421 (1992)
[21] Z Qu, G Hu and B Ma, *Phys. Lett.* **A178**, 265 (1993)
[22] K Pyragas, *Phys. Lett.* **A181**, 203 (1993)
[23] A Kittle, K Pyragas and R Richtor, *Phys. Rev.* **E50**, 262 (1994)
[24] M A Matias and J Güémez, *Phys. Lett.* **A181**, 29 (1993)
[25] M A Matias and J Güémez, *Phys. Rev. Lett.* **72**, 1455 (1994)
[26] B Peng, S K Scott and K Showalter, *J. Phys. Chem.* **94**, 1134 (1991)
[27] J Singer, Y Z Wang and H H Bau, *Phys. Rev. Lett.* **66**, 1123 (1993)
[28] Y H Chen and M Y Chou, *Phys. Rev.* **E50**, 2331 (1994)
[29] S Sinha, R Ramaswamy and J S Rao, *Physica* **D43**, 118 (1990)
[30] J K John and R E Amritkar, *Phys. Rev.* **E49**, 4843 (1994)
[31] E A Jackson, *Phys. Rev.* **A44**, 4839 (1991)
[32] R Lima and M Pettini, *Phys. Rev.* **A41**, 726 (1990)
[33] Y Braiman and I Goldhirsh, *Phys. Rev. Lett.* **66**, 2545 (1991)
[34] G Chen and X Dong, *Int. J. Bifur. Chaos* **2**, 407 (1992)
[35] S Rajasekar and M Lakshmanan, *Physica* **D67**, 282 (1993)
[36] S Rajasekar, *Chaos, Solitons and Fractals* **5**, 2135 (1995)
[37] S Parthasarthy and S Sinha, *Phys. Rev.* **E51**, 6239 (1995)
[38] K Chakrabarty and S Banerjee, *Phys. Lett.* **A200**, 115 (1995)
[39] L M Pecora and T L Carroll, *Phys. Rev. Lett.* **64**, 821 (1990); *Phys. Rev.* **A44**, 2374 (1991)
[40] T L Carroll and L M Pecora, *IEEE Trans. Circuits Syst.* **38**, 453 (1991); **40**, R1620 (1993)
[41] R E Amritkar and N Gupte, *Phys. Rev.* **A49**, R3403 (1991)
[42] L M Pecora and T L Carroll, *Phys. Rev. Lett.* **67**, 945 (1991); *Phys. Rev.* **E48**, 2426 (1993)
[43] T Kapitaniak, *Phys. Rev.* **E50**, 1642 (1994)
[44] K Murali and M Lakshmanan, *Phys. Rev.* **E49**, 4882 (1994)
[45] M Ding and E Ott, *Phys. Rev.* **E149**, R495 (1994)
[46] K M Cuomo and A V Oppenheim, *Phys. Rev. Lett.* **71**, 65 (1993)
[47] N Gupte and R E Amritkar, *Phys. Rev.* **E48**, R1620 (1993)
[48] S D Gadre and V S Varma, *Pramana – J. Phys.* **43**, 431 (1994)
[49] S D Gadre and V S Varma, *Pramana – J. Phys.* **45**, 355 (1995)
[50] E N Lorenz, *J. Atoms. Sci.* **20**, 130 (1976)
[51] M Hénon, *Commun. Math. Phys.* **50**, 69 (1976)
[52] C Sparrow, *The Lorenz equations, bifurcations and chaos* (Springer Verlag, New York, 1982)
[53] T Bohr and O B Christensen, *Phys. Rev. Lett.* **63**, 2161 (1989)
[54] K Kaneko, *Prog. Theor. Phys.* **74**, 1033 (1985); *Physica* (Amsterdam) **D23**, 437 (1986); **D34**, 1 (1989); **D37**, 60 (1989)
[55] J P Crutchfield and K Kaneko, in *Directions in chaos* edited by Hao Bai-Lin (World Scientific, Singapore, 1987); *Phys. Rev. Lett.* **60**, 2715 (1988)
[56] J D Keeler and J D Farmer, *Physica* (Amsterdam) **D23**, 413 (1986)
[57] E J Ding and Y N Lu, *J. Phys.* **A25**, 2897 (1992)
[58] G Hu and Z Qu, *Phys. Rev. Lett.* **72**, 68 (1994)

physics pp. 271–285

Effect of noise on coupled chaotic systems

MANOJIT ROY and R E AMRITKAR
Department of Physics, University of Pune, Pune 411 007, India

Abstract. The effect of noise in inducing order on various chaotically evolving systems is reviewed, with special emphasis on systems consisting of coupled chaotic elements. In many situations it is observed that the uncoupled elements when driven by identical noise, show synchronization phenomena where chaotic trajectories exponentially converge towards a single noisy trajectory, independent of the initial conditions. In a random neural network, with infinite range coupling, chaos is suppressed due to noise and the system evolves towards a fixed point. Spatiotemporal stochastic resonance phenomenon has been observed in a square array of coupled threshold devices where a temporal characteristic of the system resonates at a given noise strength. In a chaotically evolving coupled map lattice with the logistic map as local dynamics and driven by identical noise at each site, we report that the number of *structures* (a structure is a group of neighbouring lattice sites for values of the variable follow which the certain predefined pattern) follows a power-law decay with the length of the structure. An interesting phenomenon, which we call *stochastic coherence*, is also reported in which the abundance and lifetimes of these structures show characteristic peaks at some intermediate noise strength.

Keywords. Chaos; noise; coupled map lattice.

PACS Nos 05.45; 47.52

1. Introduction

Chaos in natural and human-made systems is a well established fact by now. Systems in diverse disciplines such as population biology [1], physiology [2], hydrodynamics [3], chemical reactions [4], plasma [5], lasers [6], electronics [7], computing networks [8], economic theory [9], social theory etc. have been observed to exhibit rich and unpredictable behaviour of chaos. This behaviour has been identified with the inherent nonlinear nature of the systems rather than external influences.

On the other hand, the very same nonlinearity has been seen to give rise to ordering phenomena [10], e.g., regular formation of cloud patterns, a variety of patterns in hydrodynamic flow, oscillatory patterns in chemical reactions, in the behaviour of lasers, pulse propagation in Gunn diodes etc. These patterns may be spatial, temporal, or spatiotemporal in nature and their understanding is of very special interest. One particularly interesting ordering phenomenon is the generic existence of different *structures* in a turbulent fluid [11–18]. They originate and degenerate randomly in space and time. These structures appear in spite of the fact that the fluid is undergoing a turbulent evolution and no clear understanding of this phenomenon is as yet achieved.

Noise has been known to play a detrimental role in many experimental situations. This motivated researchers to develop better techniques and methods to minimize, if not totally remove, the effect of noise and enhance the signal-to-noise ratio and hence system performance. A slight amount of noisy perturbation is known to destroy delicate patterns in spatially extended systems.

In the light of these facts, interest grew when evidence to the contrary started appearing as regards the effect of noise on ordering phenomena. Firstly, it has been observed that addition of noise of a given strength to certain nonlinear systems increases the system's response at a particular time scale, thereby improving the signal-to-noise ratio; these findings opened up an entirely new field of research, known as *stochastic resonance* [19–27]. Secondly, noise has been seen to influence spatial and spatiotemporal behaviour of some nonlinear systems in a quite counterintuitive manner. It is observed that noise can trigger, select and sustain patterns in optical systems, fluid dynamical systems, etc. [28–32].

Considerable progress has been made to establish a connection between these pattern forming systems (such as fluids) and the theory of nonlinear dynamical systems. A natural next step is to see how noise influences the spatiotemporal evolution of these dynamical systems, with the hope that this kind of study may shed some light on the abovementioned noise-induced features in physical systems. This article presents a review of the work that has been carried out on the effect of noise in inducing order in an otherwise chaotically evolving system, with particular emphasis on systems consisting of coupled chaotically evolving elements.

We have organized the article as follows: Section 2 discusses noise induced ordering phenomena observed in low-dimensional uncoupled systems. In § 3, we review the work on effect of noise on two spatially extended systems, a neural network and array of firing elements. Section 4 introduces the coupled map lattice (CML), one of the most popular models of nonlinear dynamical systems with spatial extension. In this same section we discuss a recent work that we have carried out on the effect of noise on CML. Section 5 summarizes and concludes the article.

2. Effect of noise on uncoupled systems

In this section we investigate some examples of noise-induced ordering in uncoupled systems.

2.1 *Synchronization*

In some systems different trajectories get synchronized asymptotically to a single noisy trajectory independent of the initial conditions when driven by an identical sequence of noise above a certain strength. Here, synchronization should be understood as the exponential convergence of the average distance between any two phase space points. This means that synchronization is essentially a nonchaotic phenomenon associated with a negative Lyapunov exponent [33], although the asymptotic trajectory can be very random. (The Lyapunov exponent λ characterizes the rate at which the distance between neighbouring trajectories changes. If ϵ_0 is the initial separation between two trajectories, the separation after time t can be written as $\epsilon_t \simeq \epsilon_0 \exp(\lambda t)$. A positive λ implies exponential divergence of nearby trajectories and hence chaos. For multidimensional systems our reference to 'Lyapunov exponent' will always mean the largest of the Lyapunov characteristic exponents.)

In the continuum-time limit such systems can be written as a set of uncoupled Langevin equations

$$\dot{x}_i(t) = F(x_i) + \eta(t), \tag{1}$$

where x_i is a dynamical variable, the index i corresponds to different trajectories, i.e., different initial conditions, the (nonlinear) function $F(x)$ governs the dynamics of the system, and $\eta(t)$ is a delta-correlated ($\langle\eta(t)\eta(t')\rangle = \langle\eta^2(t)\rangle\delta(t-t')$) noisy driving force imparted at a regular interval that is small compared to all relevant macroscopic time scales of the system, and is the same for all i. Equations (1) have the synchronization solution $x(t) = x_i(t)$. Therefore, the question essentially is whether this solution is stable or not for a given range of system parameters. In other words, if an appropriately constructed $\lambda < 0$, there is an exponential convergence of trajectories and hence synchronization, and if $\lambda > 0$, the evolution is chaotic.

Let us now take up individual cases.

2.1.1 *Particle in a Newtonian potential*: Fahy and Hamann (FH) [34] have studied a Newtonian particle moving in a smooth multiminima potential $V(x)$ without friction, subjected to the condition that at regular time intervals τ, it is stopped and all its velocity components are reset to random values chosen from a Gaussian distribution. They observed that when an ensemble of such particles with different initial conditions is driven by an identical sequence of random forces, their trajectories asymptotically get synchronized to a single noisy trajectory provided that τ is less than certain critical value τ_c. On the basis of this observation they concluded that for $\tau < \tau_c$ the final trajectory of the particles is independent of the initial conditions to any required level of accuracy (the accuracy aspect will have importance later in our discussion); it depends only on the choice of velocities. Thus the trajectory, albeit noisy, is not chaotic for $\tau < \tau_c$ because of exponential convergence, and an appropriately defined Lyapunov exponent λ is negative. They have shown that for any one-dimensional potential $V(x)$ confining the particles to a finite region and for sufficiently small τ, the average rate of contraction γ of the distance between two particles initially close together is given by $\gamma = \tau\beta\langle(\partial V/\partial x)^2\rangle/2m + O(\tau^2)$, where $\beta = 1/k_BT$, and m is the mass of the particle (angular brackets denote an average with respect to the Boltzmann distribution). The above mentioned synchronization phenomenon is a stronger observation than the well-known statistical phenomenon in which the asymptotic distribution of Brownian trajectories is found to be proportional to the Boltzmann factor $\exp[-\beta V(x)]$, independent of their initial distribution. FH have also conjectured that this synchronization feature may be generic to all bounded systems.

2.1.2 *Noisy chaotic systems*: Maritan and Banavar (MB) [35] have considered the effect of noise on the following two chaotic systems. First, they have taken a noisy logistic map with the evolution law

$$x_{t+1} = 4x_t(1-x_t) + \eta_t, \tag{2}$$

where η_t is a uniform-deviate random number chosen from an interval $[-W, +W]$ with the constraint that $0 < x_{t+1} < 1$. Equation (2) may be considered as an example of the discrete-time version of eq. (1) with $\Delta t = 1$. They found a critical $W_c(\approx 0.5)$ such that when a pair of randomly chosen initial conditions is driven by an identical sequence of η_t with $W > W_c$, their asymptotic trajectories become synchronized (within a given accuracy) to a single random trajectory independent of the initial conditions. They have computed the mean squared separation $\bar{d}^2$ between the two identically driven trajectories and observed that $\bar{d}^2$ falls off for $W > W_c$. They offered an intuitive explanation saying that the convergence can occur for logistic map due to the contraction of d whenever the sum of the pair of numbers comes close to unity, because the distance evolves as

$$|x_{t+1}(1) - x_{t+1}(2)| = 4|x_t(1) - x_t(2)|[1 - \{x_t(1) + x_t(2)\}].$$

Since peaks in the invariant density for the logistic map are close to 0 and 1, $W \geq \frac{1}{2}$ can bring both $x(1)$ and $x(2)$ near $\frac{1}{2}$.

The second system considered by MB is the Lorenz system described by

$$\frac{dx}{dt} = \sigma(x - y), \quad \frac{dy}{dt} = -xz + rx - y, \quad \frac{dz}{dt} = xy - bz, \tag{3}$$

with $\sigma = 10$, $b = \frac{8}{3}$, and $r = 28$. The y-equation is then evolved in difference form as

$$y(t + \Delta t) = y(t) + [-x(t)z(t) + rx(t) - y(t)]\Delta t + \eta_t W_l \sqrt{\Delta t}, \tag{4}$$

where η_t is again a delta-correlated uniform-deviate random number between $[0, 1]$ (it has a nonzero mean unlike in the logistic case) and W_l is the amplitude. Note that here addition of noise is unrestricted unlike in the logistic case where the boundedness of the phase space constrains the noise to depend on the state of the system. For the Lorenz system also a threshold value $(\approx \frac{2}{3})$ for W_l was found beyond which the synchronization phenomenon was observed for an identically driven system with different initial conditions. It is reported that synchronization does not occur if η_t with the same amplitude has zero mean.

MB have maintained that although the systems considered above exhibit synchronization, their strange attractors are not replaced by topologically simple structures like a fixed point or limit cycle.

Pikovsky made a cautionary observation [36] that for all bounded systems there is always a nonzero probability that any two phase space points will come close to within some $\epsilon > 0$ with or without noise. In other words after a sufficiently long time the two systems will be synchronized because of the finite precision (of the computer) and finiteness of phase space. This type of spurious synchronization will occur even for systems with positive λ, which is to be distinguished from the physical synchronization which is characterized by negative λ. The former is unstable against small perturbations because of positive λ and so will not be observed in real experiments, whereas the latter is quite stable. Pikovsky found λ for the noisy logistic system of MB to be positive and concluded that this type of synchronization is a numerical artifact.

MB subsequently noted [37] that whether or not a physical synchronization occurs, identical noise with sufficient strength drastically enhances the probability of a 'close encounter' between any two ensemble points.

There has been an attempt [38] to interpret synchronization in terms of the inherent structural instability of the undriven system. These authors maintain that the strange attractor of the undriven chaotic system gets replaced by a stable fixed point under parametric perturbations, resulting in synchronization of any two phase space points and a negative λ. They showed this for all the three systems mentioned above by treating noise as a perturbation in the parameter and also considered new examples. It was remarked that this phenomenon is not generic and holds for only those systems in which this type of structural instability occurs.

The picture that is emerging out of these findings is that while an understanding of the synchronization phenomenon may still remain incomplete, there is no doubt about noise playing a crucial role in drastically increasing the chance of a close encounter of any two phase space points and thereby bringing down the Lyapunov exponents of the otherwise chaotic system.

2.2 *Violation of the law of large numbers*

An interesting observation was reported [39] regarding the chaotic evolution of an ensemble of uncoupled maps driven by a parametric noise. The system consists of N local maps

$$x_{t+1}(i) = F(x_t(i);\ a_t(i)), \tag{5}$$

where the nonlinearity parameter $a_t(i)$ is subjected to fluctuations in both space and time. The quantity of interest is the mean field h_t, defined as

$$h_t = \frac{1}{N}\sum_{j=1}^{N} F(x_t(j)).$$

For uncoupled variables $x_t(i)$ fluctuating almost independently, h_t for large N is expected to obey the law of large numbers, and hence the mean-square deviation (MSD) $(= \langle h_t^2\rangle - \langle h_t\rangle^2)$ should vary as $1/N$, and converge to a fixed-point value as $N \to \infty$. It was observed that with the logistic map in the chaotic regime, the MSD falls off as $\sim 1/N$ if $a_t(i)$ fluctuates in space. On the other hand if $a_t(i) = a_t$ is independent of i then the MSD saturates with N beyond a critical $N = N_c$, whose value depends on the strength of the noise. Similar behaviour was observed using other maps such as the circle and tent maps. The author claims that an ensemble of uncoupled chaotic maps with spatially uniform parametric fluctuations violates the law of large numbers, irrespective of the details of the map.

3. Effect of noise on spatially extended systems

We shall now consider systems which have a spatial extension. Because of increased complexity owing to the largely enhanced phase space dimensions (one mostly talks of infinite dimensional phase spaces) these systems show very rich dynamical features. Spatially extended systems are modelled by a set of variables (whose evolution may be governed by discrete-time nonlinear maps like the logistic map, or continuous-time ordinary differential equations such as the oscillator equation) physically coupled to each other in euclidean space.

Let us consider a few cases of such systems and the influence of noise on their evolution.

3.1 *Noise and suppression of chaos in a neural network*

Molgedey, Schuchhardt and Schuster (MSS) have investigated the effect of noise on discrete-time evolution of a random neural network with infinite-range interactions [40]. The model consists of N analog neurons $\{x_t(i)\}$, $i = 1, \ldots, N$, with $-1 \leq x(i) \leq 1$, evolving according to the law

$$x_{t+1}(i) = F(h_t(i)), \tag{6}$$

$$h_t(i) = \sum_{j \neq i} \varepsilon_{ij} x_t(j) + \eta_t(i), \tag{7}$$

where the function $F(h)$ has the following properties: 1. it is odd $[F(-h) = -F(h)]$; 2. it approaches ± 1 as $h \longrightarrow \pm\infty$; 3. it increases near $h = 0$ as $[\mathrm{d}F/\mathrm{d}h]_{h=0} = g$, where $g > 0$ is the gain parameter. The coupling parameters ε_{ij} are delta-correlated Gaussian random variables with zero mean, $h_t(i)$ denotes the internal field of the neuron and $\eta_t(i)$ is the external white noise with zero mean and variance $\langle \eta_t(i)\eta_\tau(j) \rangle = \sigma^2 \delta_{ij} \delta_{t\tau}$.

MSS used a dynamical functional approach to reduce the dynamics of the entire system to an equation for an effective single neuron in the thermodynamical limit:

$$x_{t+1} = F(h_t), \tag{8}$$

with

$$\langle h_t h_\tau \rangle = \sigma^2 \delta_{t\tau} + \langle F(h_{t-1}) F(h_{\tau-1}) \rangle \tag{9}$$

Equation (8) with eq. (9) yields the same averaged dynamical properties as for eq. (6). One now defines the activity of the network as $K_t = \langle h_t^2 \rangle$ which for small K reduces to

$$K_t = \sigma^2 + g^2 K_{t-1} + O(K_{t-1}^2). \tag{10}$$

In the absence of noise there is a trivial fixed point $K^* = 0$ for eq. (10) and is stable for $g < 1$. In the presence of noise it has only one stable fixed point in the range $[0, 1 + \sigma^2]$.

To study the chaotic behaviour of the network a replica of the system with infinitesimally separated initial conditions and with the same noise $\eta_t^1(i) = \eta_t^2(i)$ is constructed. The Lyapunov exponent for the system is defined as

$$\lambda = \lim_{\tau \rightarrow \infty} \lim_{x_t^1 \rightarrow x_t^2} \frac{1}{2\tau} \log_2 \frac{\langle (x_{t+\tau}^1 - x_{t+\tau}^2)^2 \rangle}{\langle (x_t^1 - x_t^2)^2 \rangle}. \tag{11}$$

Assuming equilibrium ($\langle (h_t^1)^2 \rangle = \langle (h_t^2)^2 \rangle$), they [40] have obtained for the noiseless case $\lambda < 0$ for $g \leq 1$ and the system goes to the trivial fixed point $K^* = 0$. For $g > 1$, $\lambda > 0$ and the system shows chaos.

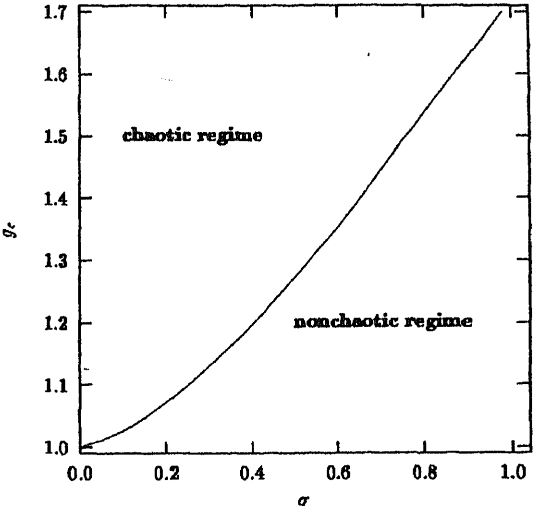


Figure 1. Plot of phase diagram (g_c versus σ) obtained by Molgedey *et al* [40] for an infinite range neural network with external white noise. For a given gain g chaos is suppressed for sufficiently strong noise σ.

They have numerically studied the effect of noise on the chaotic properties of the system using the following form for $F(h)$:

$$F(h) = \begin{cases} -1 & \text{for } h < -1/g, \\ gh & \text{for} - 1/g \leq h \leq +1/g, \\ +1 & \text{for} + 1/g < h. \end{cases}$$

For g less than some critical value g_c, the activity K^* increases with noise as the system settles onto the stable fixed point $K^* = 1 + \sigma^2$, whereas for $g > g_c$ chaos sets in. It was found that for small noise asymptotically $g_c = 1 - \sigma^2 \ln \sigma^2$ and for large noise $g_c = \sqrt{\pi/2}\,\sigma$. Figure 1 is the plot of g_c versus σ (the phase diagram), which clearly shows the chaotic and regular regimes.

MSS concluded that for higher dimensional systems of this type where chaos occurs essentially due to the randomized interactions amongst constituent nonchaotic elements, noise acts to impair information flow between these elements and thereby suppresses chaos.

3.2 *Spatiotemporal stochastic resonance in excitable media*

Let us now consider the observation of the stochastic resonance phenomenon in a spatially extended pattern forming system, as reported by Jung and Mayer–Kress (JMK) [41]. Stochastic resonance, as the name suggests, is a phenomenon in which at a given noise-strength certain temporal patterns of the system get enhanced drastically owing to an increase in the system's sensitivity [19, 20, 25]. This behaviour has a characteristic bell-shaped curve when the output power (of the Fourier spectrum) at the corresponding frequency is plotted against noise-strength, with the peak at the 'resonating' noise value.

two-dimensional equidistant square array (with lattice constant a) of $N \times N$ noisy threshold devices $d(i,j), i,j = 1,\ldots,N$ with the following properties: if the input $x_t(i,j)$ to the devices is below some threshold b, output $y_t(i,j) = 0$; if $x_t(i,j) > b$ the device generates a spiky output of the form $y_t(i,j) = I_0\Theta(\dot{x}(i,j))\,\delta(x_t(i,j) - b)$ and then goes into temporary hibernation for a refractory period Δt_r. The dynamics can be described by a linear Langevin equation for inputs $x_t(i,j)$ as

$$\dot{x}(i,j) = -\gamma x(i,j) + \sqrt{\gamma\sigma}\,\eta_t(i,j), \tag{12}$$

where γ is a leakage constant which accounts for the thermal leakage due to interaction with the surroundings, $\eta_t(i,j)$ is a delta-correlated noise and $\sigma = \langle x_t(i,j)^2\rangle$. Equation (12) on integration over Δt gives rise to the discrete time dynamics

$$x_{t+\Delta t}(i,j) = x_t(i,j)\exp(-\gamma\Delta t) + G_t(i,j), \tag{13}$$

where $G_t(i,j)$ is a Gaussian random number with variance $\sigma_G = \sigma[1 - e^{-2\gamma\Delta t}]$.

Threshold devices are pulse-coupled, i.e., when an element $d(k,l)$ fires, it communicates with its surrounding elements $d(i,j)$ at a distance $r_{ij,kl}$ by contributing an amount $K\exp(-\beta r^2_{ij,kl}/a^2)$ to their inputs $x_t(i,j)$, where $K \to K/b$ is in dimensionless unit and β is a dimensionless quantity describing the spatial coupling range. For large coupling ($K \geq \exp(\beta)$), these firing elements give rise to excitatory waves spreading through the array, such as spiral waves, target waves (single, nonrepetitive wavefronts), etc. The selection of the wave form depends on the geometry of initial conditions. There is a threshold K_c such that for $K < K_c$ excitatory waves do not necessarily start.

However, in the presence of noise the spreading of excitatory waves, such as spiral waves, is observed in subthreshold regime ($K < K_c$). Any such noise-sustained wave

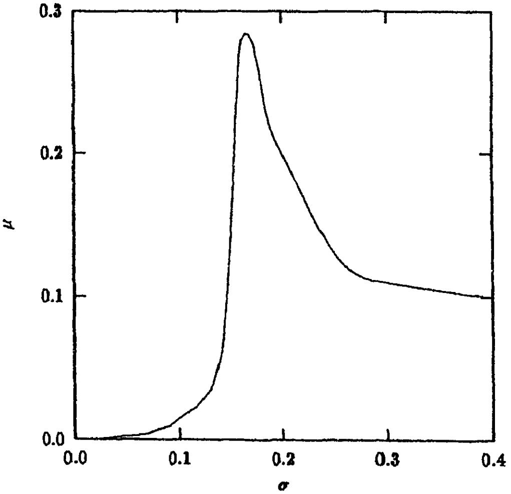


Figure 2. Plot of the graph of the time averaged number of excess events μ with noise variance σ, for a typical lattice size 200 × 200 and with a given set of parameter values [41].

disappears soon after the noise is turned off. With increasing noise the spiral wave evolves with a larger curvature, i.e., with an enhanced coherence in the array. If the noise is increased further, the spiral breaks up and there is a loss of coherence.

In order to show that the phenomenon observed above is essentially spatiotemporal stochastic resonance, JMK have taken a solitary initial wave and defined a quantity μ, the time averaged number of excess events, as the difference between the number of firing elements along the driving wavefront and the average number of firing events along a row of the array, not affected by the driving. The plot of μ versus σ, the variance of the noise, is shown in figure 2. The plot exhibits the characteristic bell-shaped curve mentioned earlier. Synchronization of the spatiotemporal firing pattern to an external driving, thereby giving rise to the resonating peak, is a simple generalization of the stochastic resonance to this extended system.

We shall now move on to the final part of the paper, namely, the effect of noise on coupled map lattices (CML). As the model itself has a plethora of interesting dynamical features, we begin with an introduction to CML.

4. Coupled map lattices

Coupled map lattices (CML) are spatially extended dynamical systems with discrete space (lattice), discrete time and continuous state evolution. General evolutionary dynamics of a CML may be expressed as

$$x_{t+1}(i) = F(x_t(i)) + \sum_{j\neq i} \varepsilon_{ij} G(x_t(j), x_t(i)), \tag{14}$$

where F and G are nonlinear maps and the state variables x_i vary continuously in the phase space of the map, and i is the space index. CMLs are in general constructed with the symmetry properties of spatial translational and rotational invariance, and so the coupling term ε_{ij} is taken to be a uniform scalar ranging from the nearest-neighbour coupling to global coupling.

CMLs have been used to model many physical phenomena. These include pattern formation, chemical waves, excitable media, nucleation, crystal growth, charge density waves, population dynamics, fluid dynamics, etc. [42–55]. Chemical systems exhibit very rich spatiotemporal structures and patterns. The standard description of these systems is in terms of the reaction-diffusion equations that incorporate the combined effects of local nonlinear reaction dynamics and global diffusion of chemical species due to the concentration gradients in the system. Many of the averaged features and even some detailed ones of chemical patterns can be captured by a simple CML description of the dynamics. Another area of interest is the phase ordering dynamics. Phase separation occurs due to competition between different states in the system, such as domain growth in ferromagnetic and chemical systems. Some of the salient features of phase ordering dynamics are also observed in the CML model. CML may also be applied in pattern dynamics with an excitable state and a relaxation from that state. Such processes are observed in reaction-diffusion in excitable media and also in biological problems such as heart rhythm and electrical activities in neural tissues. In fluid dynamics, simulations using CML can show formation of convective patterns, vortices, sinks, behaviour. One of the major interests in studying CML is in the context of understanding spatial and temporal structure formation, specially in fluid dynamics. These structures appear in spite of the fact that the underlying evolution of the system is spatiotemporally chaotic.

One of the most extensively studied types of one dimensional CML is that with nearest-neighbour coupling and having the following form:

$$x_{t+1}(i) = F(x_t(i)) + \frac{\varepsilon}{2}\left[G(x_t(i-1)) + G(x_t(i+1)) - 2G(x_t(i))\right]. \tag{15}$$

We shall concentrate on the form $G(x) = F(x)$, which reduces eq. (15) to

$$x_{t+1}(i) = (1-\varepsilon)F(x_t(i)) + \frac{\varepsilon}{2}[F(x_t(i-1)) + F(x_t(i+1))]. \tag{16}$$

This particular model is known as the *future diffusive* CML since the diffusively coupled entities are one timestep evolved values. Studies of this model alongwith the local dynamics as the logistic map $F(x) = \mu x(1-x)$ have revealed that they can exhibit a wide range of spatiotemporal complexity. It has been observed that the temporal period doubling behaviour of the map can induce spatial domain structures separated by kinks and antikinks. A pattern selection regime is observed where patterns of certain characteristic lengths are selected. Phenomena such as spatiotemporal quasiperiodicity, soliton turbulence, spatiotemporal intermittency, wavelength doubling bifurcations, synchronization etc. for this and other maps (like the circle map $\theta_{t+1} = \theta_t + \Omega - (K/2\pi)\sin(2\pi\theta_t)$, Ω-angular frequency, K-nonlinearity parameter) have been identified [56–64].

We now discuss the effect of noise on the evolution of a CML. In the next subsection we report a novel phenomenon that we have observed in this context [65].

4.1 *Noise and 'stochastic coherence' in a CML*

We take the system of eq. (16) with an additional noise term as follows:

$$x_{t+1}(i) = (1-\varepsilon)F(x_t(i)) + \frac{\varepsilon}{2}[F(x_t(i-1)) + F(x_t(i+1))] + \eta_t, \tag{17}$$

where η_t is the familiar delta-correlated noise. The logistic function $F(x) = \mu x(1-x)$ is used as local dynamics. For noise η_t we have used a uniformly distributed random number bounded between $-W$ and $+W$, with the constraint that $0 < x_{t+1}(i) < 1$; we call W the noise-strength parameter. Values of μ, ε, and L (size of the lattice) are chosen so that the resultant dynamics of the system is chaotic.

Now, we define a *structure* as a region of space such that the difference in the values of the variables of neighbouring sites within this region is less than a predefined small positive number, say δ, i.e., $|x_t(i) - x_t(i \pm 1)| \leq \delta$. We call δ the structure parameter. We look for such patterns, or structures, to appear in the course of evolution of the model given by eq. (17).

Figure 3 shows a plot (on a log–log scale) of the distribution of the number $n(l)$ vs. length l of the structures for different values of W, with $\mu = 4.0$, $\epsilon = 0.6$, $\delta = 0.0001$ and $L = 1000$, and open-boundary conditions. The power-law nature of the decay of $n(l)$ is

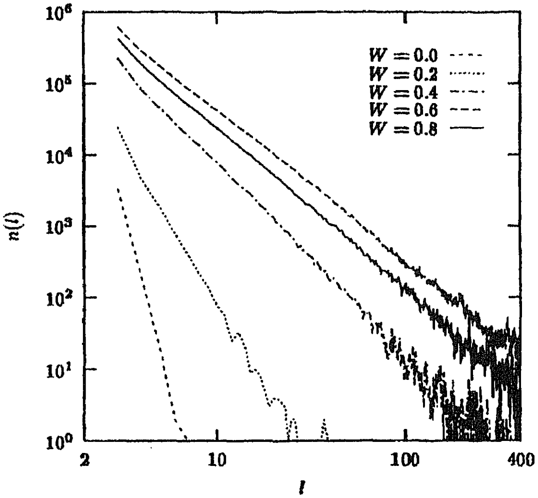


Figure 3. Plot of variation of number $n(l)$ of structures with length l for a lattice with size $L = 1000$, for different values of noise-strength W as indicated. Parameters chosen are $\varepsilon = 0.6$, $\delta = 0.0001$, and $\mu = 4.0$. Open-boundary conditions are used. Data are obtained for 50000 iterates, each for 10 initial conditions.

clearly evident, which has a form

$$n(l) \propto l^{-\alpha}, \tag{18}$$

where α is the power-law exponent. This indicates that the system does not have any intrinsic length scale. It may be noted that in the absence of noise ($W = 0.0$) the decay is manifestly exponential [66]. The exponent α is seen to depend on the noise-strength W, with a minimum for W around 0.6. We define the average length $\bar{l}$ of a structure as $\bar{l} = \sum l\, n(l) / \sum n(l)$. In figure 4 we plot the variation of $\bar{l}$ with W for values of parameters as in figure 3. The plot exhibits a *bell-shaped* nature within a fairly narrow range of W around the value 0.6 (one may note the surprising similarity between figures 4 and 2, though they refer to completely different phenomena). It may be noted that the minimum of α also occurs for W quite close to this value, as expected.

We call the phenomenon observed above *stochastic coherence*. This is similar to stochastic resonance which shows a bell-shaped behaviour of the temporal response as a function of the noise-strength, as mentioned before. However, one may note that our system does not have any intrinsic length-scale, whereas in stochastic resonance the noise resonates with a given time-scale; hence our use of the word coherence rather than resonance. In stochastic resonance noise transfers energy to the system at a characteristic frequency, whereas in stochastic coherence noise induces coherence in the system.

To study the evolutionary aspects of these structures we obtained the distribution of the number $n(\tau)$ of structures vs. their lifetimes τ for different W. $n(\tau)$ is observed to decrease with τ with a stretched exponential type of decay having a form

$$n(\tau) \propto \exp(-(\text{const.})\tau^{\zeta}), \tag{19}$$

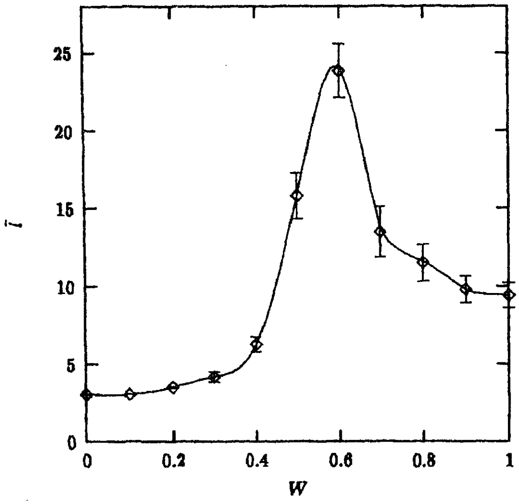


Figure 4. Plot of variation of average length $\bar{l}$ of structure with noise-strength W, with parameters as stated in figure 3.

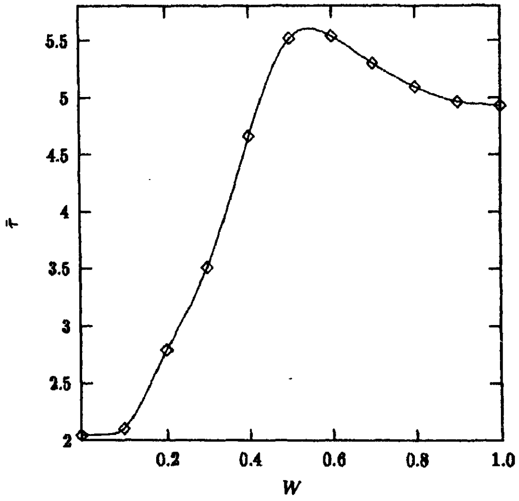


Figure 5. Variation of average lifetime $\bar{\tau}$ of structures with noise-strength W shown for parameters as stated in figure 3.

where ζ depends on W. We define the average lifetime $\bar{\tau}$ of a structure as $\bar{\tau} = \sum \tau n(\tau) / \sum n(\tau)$. In figure 5 we plot $\bar{\tau}$ vs. W for parameter values as in figure 3. The graph shows a bell-shaped feature with a maximum for W around 0.6.

In order to ascertain the chaotic nature of the system evolution we have calculated the Lyapunov exponent spectra for our system. We find a number of Lyapunov exponents to be positive, implying that the underlying evolution is chaotic. The maximum Lyapunov exponent λ shows a minimum around noise-strength 0.6. The fact that we have observed other extrema for similar W may make it appear that the origin of these behaviours probably lies in the reduction of λ due to noise [67] (though it should be noted that the reduction is not monotonic). To explore this possibility we have studied the variation of λ with the coupling parameter ε. We find that λ remains fairly constant for $0.2 \leq \varepsilon \leq 0.8$ for the entire range of W. On the other hand, a plot of the variation of the average length $\bar{l}$ with ε for fixed W shows a monotonically increasing behaviour, quite contrary to what is expected from λ. This implies that the Lyapunov exponent alone cannot be used for proper characterization of the spatio-temporal features of the system (unlike the synchronization phenomenon discussed earlier, where λ is sufficient to fully characterize the behaviour).

To conclude this section, we have reported a new phenomenon observed in a chaotically evolving one-dimensional CML driven by an identical noise, which we termed stochastic coherence. It is observed that there is a phenomenal increase in the abundance of coherent structures of all scales due to noise. The distribution of these structures shows a power-law decay with the length of the structure. The average length as well as the average lifetime of these structures exhibit characteristic maxima at certain noise-strength values.

5. Summary and conclusion

We have reviewed the work that has been carried out on the effect of noise on evolutionary dynamics of chaotic systems. In the case of uncoupled systems, we have encountered a synchronization phenomenon for different phase space trajectories driven by identical noise above a certain strength. We have also seen nonstatistical behaviour for an ensemble of chaotic trajectories driven by an identical sequence of parametric noise. For coupled systems, noise above a given strength has been observed to suppress chaos in a random neural network with infinite range interactions. Spatiotemporal stochastic resonance, the first case of a stochastic resonance phenomenon observed in systems with spatial extension, has been reported for a two-dimensional square array of firing devices. In the context of CML, we have observed an interesting phenomenon, stochastic coherence, when the entire lattice is driven by identical noise.

As noted earlier in the text, understanding the formation and evolution of patterns and structures in spatially extended systems is still incomplete. In real physical situations, noise almost ubiquitously influences the system behaviour. So while studying the ordering phenomena, incorporating noise as a part of the evolving 'supersystem' and studying the mutual interaction of the two subsystems offers a more practical way of looking into the problem. The inherent nonlinearity in the system, which on the one hand gives rise to chaos and on the other hand ordered behaviour, is surely again playing its role in enhancing order under the influence of noise; how so is still not quite clear. We have encountered two kinds of noise-induced ordering phenomena: global ordering (such as synchronization) and local ordering (spatiotemporal stochastic resonance and stochastic coherence). Stochastic coherence, in particular, may turn out to be a promising way towards achieving any degree of understanding of the structure formation in extended systems, although a lot of work remains to be done. While it is definitely not claimed that all questions will be answered with this approach, we certainly hope to bridge a few gaps in the existing knowledge.

Acknowledgments

One of the authors (MR) acknowledges the University Grants Commission (India) and the other author (REA) acknowledges the Department of Science and Technology (India) for financial assistance.

References

[1] R M May, *Nature (London)* **261**, 459 (1976)
[2] M C Mackey and L Glass, *Science* **197**, 287 (1977)
[3] A Libchaber and J Maurer, in *Nonlinear phenomena at phase transitions and instabilities* edited by T Riste (1982) p. 259
[4] S K Scott, *Chemical Chaos* (Oxford University Press, Walton Street, Oxford, 1993)
[5] P Y Cheumg and A Y Wong, *Phys. Rev. Lett.* **59**, 551 (1987)
[6] J R Tredicce and L M Narducci, in *Directions in chaos* edited by Hao Bai Lin (World Scientific, Singapore, 1987) vol. 1, p. 148
[7] P S Linsay, *Phys. Rev. Lett.* **47**, 1349 (1981)
[8] J O Kephart, T Hogg and B A Huberman, *Phys. Rev.* **A40**, 404 (1989)
[9] T Hogg, B A Huberman and M Youssefmir, preprint (1995)
[10] *Order and chaos in nonlinear physical systems* edited by S Lundqvist, N H March and M P Tosi (Plenum Press, New York, 1988)
[11] R W Schefer *et al, Phys. Fluids* **6**, 652 (1994)
[12] K H Park, R W Metcalf and F Hussain, *Phys. Fluids* **6**, 885 (1994)
[13] E Stone and P Holmes, *Physica* **D37**, 20 (1989)
[14] S D Meyeres, J Sommeria and H L Swinney, *Physica* **D37**, 515 (1989)
[15] A K M F Hussain, *Phys. Fluids* **26**, 2816 (1983)
[16] L Sirovich, *Physica* **D37**, 126 (1989)
[17] *Topological fluid mechanics* edited by H K Moffatt *et al*, Proceedings of the IUTAM symposium 1989, Cambridge University Press (1990)
[18] *Turbulence and stochastic processes: Kolmogorov's ideas 50 years on* edited by J C R Hunt *et al* (The Royal Society, London, 1991)
[19] R Benzi, A Sutera and A Vulpiani, *J. Phys.* **A14**, 1453 (1981)
[20] B McNamara and K Weisenfeld, *Phys. Rev.* **A39**, 4854 (1989)
[21] L Gammaitoni, F Marchesoni, E Manichella–Saetta and S Santucci, *Phys. Rev. Lett.* **62**, 349 (1989)
[22] Hu Gang, G Nicolis and C Nicolis, *Phys. Rev.* **A42**, 2030 (1990)
[23] M I Dykman, R Mannella, P V E McClintock and N G Stocks, *Phys. Rev. Lett.* **65**, 2606 (1990)
[24] *J. Stat. Phys.*, edited by F Moss, A Bulsara and M Shlesinger **70**, 1–514 (1993)
[25] R F Fox and Yan–nan Lu, *Phys. Rev.* **E48**, 3390 (1993)
[26] A R Bulsara and L Gammaitoni, *Phys. Today* **39** (1996)
[27] P Jung, *Phys. Rev.* **E50**, 2513 (1994)
[28] *Noise in dynamical systems* edited by F Moss and P V E McClintock (Cambridge University Press, Cambridge, U.K., 1989) vol. 1–3
[29] *Noise and Chaos in Nonlinear Dynamical Systems* edited by F Moss, L A Lugiato and W Schleich (Cambridge University Press, Cambridge, U.K., 1990)
[30] D A Kurtze, *Phys. Rev. Lett.* **77**, 63 (1996)
[31] R J Deissler, *Phys. Lett.* **A120**, 334 (1987)
[32] H R Brand and R J Deissler, *Phys. Lett.* **A130**, 293 (1988)
[33] A Pikovsky, in *Nonlinear and turbulent processes in physics* edited by R Z Sagdeev (Harwood Acad., Chur, Switzerland, 1984); *Phys. Lett.* **A165**, 33 (1992)
[34] S Fahy and D R Hamann, *Phys. Rev. Lett.* **69**, 761 (1992)
[35] A Maritan and J R Banavar, *Phys. Rev. Lett.* **72**, 1451 (1994)

[36] A Pikovsky, *Phys. Rev. Lett.* **73**, 2931 (1994)
[37] A Maritan and J R Banavar, *Phys. Rev. Lett.* **73**, 2932 (1994)
[38] P M Gade and C Basu (to be published in *Phys. Lett.* A)
[39] S Sinha, *Phys. Rev. Lett.* **69**, 3306 (1992)
[40] L Molgedey, J Schuchhardt and H G Schuster, *Phys. Rev. Lett.* **69**, 3717 (1992)
[41] P Jung and G Mayer–Kress, *Phys. Rev. Lett.* **74**, 2130 (1995)
[42] G L Oppo and R Kapral, *Phys. Rev.* **A36**, 5820 (1987)
[43] D Kessler, H Levine and W N Reynolds, *Phys. Rev.* **A42**, 6125 (1990)
[44] Y Oono and S Puri, *Phys. Rev. Lett.* **58**, 836 (1986)
[45] T Yanagita and K Kaneko, *Phys. Lett.* **A175**, 415 (1993)
[46] M H Jensen, *Phys. Rev. Lett.* **62**, 1361 (1989)
[47] R V Sole and J Valls, *Phys. Lett.* **A153**, 330 (1991)
[48] I Waller and R Kapral, *Phys. Rev.* **A30**, 2047 (1984)
[49] K Kaneko, *Prog. Theor. Phys.* **74**, 1033 (1985)
[50] R J Deissler and K Kaneko, *Phys. Lett.* **A119**, 397 (1987)
[51] F H Willeboordse and K Kaneko, *Phys. Rev. Lett.* **73**, 533 (1994)
[52] J D Keeler and J D Farmer, *Physics* **D23**, 413 (1986)
[53] S Sinha, D Biswas, M Azam and S V Lawande, *Phys. Rev.* **A46**, 6242 (1992)
[54] I Aranson, D Golomb and H Sompolinski, *Phys. Rev. Lett.* **68**, 3495 (1992)
[55] M G Cosenza and R Kapral, *Phys. Rev.* **A46**, 1850 (1992)
[56] J P Crutchfield and K Kaneko, in *Directions in chaos* (World Scientific, Singapore, 1987) vol. 1, 272
[57] *Theory and applications of coupled map lattices* edited by K Kaneko (John Wiley and Sons, Chichester, England, 1993)
[58] K Kaneko, *Phys. Lett.* **A149**, 105 (1990)
[59] H Chate and P Manneville, *Physics* **D32**, 409 (1988)
[60] R E Amritkar and P M Gade, *Phys. Rev. Lett.* **70**, 3408 (1993)
[61] R E Amritkar, P M Gade, A D Gangal and V M Nandakumaran, *Phys. Rev.* **A44**, 3407 (1991)
[62] P M Gade and R E Amritkar, *Phys. Rev.* **E47**, 143 (1993)
[63] Qu Zhilin and Hu Gang, *Phys. Rev.* **E49**, 1099 (1994)
[64] N Chatterjee and N M Gupte, *Phys. Rev.* **E53**, 4457 (1996)
[65] M Roy and R E Amritkar (communicated to *Phys. Rev.* E)
[66] J K John and R E Amritkar, *Phys. Rev.* **E51**, 5103 (1995)
[67] K Matsumoto and I Tsuda, *J. Stat. Phys.* **31**, 87 (1983); **34**, 111 (1984)

Lattice dynamical models of adaptive spatio-temporal phenomena

SUDESHNA SINHA
The Institute of Mathematical Sciences, CIT Campus, Madras 600 113, India

Abstract. We describe the rich spectrum of spatio-temporal phenomena emerging from a class of models incorporating adaptive dynamics on a lattice of nonlinear (typically chaotic) elements. The investigation is based on extensive numerical simulations which reveal many novel dynamical phases, ranging from spatio-temporal fixed points and cycles of all orders, to parameter regimes displaying marked scaling properties (as manifest in distinct $1/f$ spectral characteristics and power law distributions of spatial quantities).

Keywords. Spatio-temporal chaos; lattice dynamical models.

PACS No. **05.45**

1. Introduction

A basic aim of modelling in physics is to provide suggestive conceptual frameworks for understanding complex phenomena, which are generic in physical systems. One seeks to construct simple yet effective models which are capable of capturing the essence of complicated dynamic processes as they occur in nature. For instance, one dimensional unimodal maps can serve as a versatile paradigm for low dimensional dissipative chaos, as they yield an extraordinarily rich spectrum of behaviours. From low dimensional nonlinear systems, which are now reasonably well understood, we reach out towards *nonlinear extended* systems. The hope is to build prototypes for such composite complex systems — prototypes which can yield a repertoire of dynamical behaviour reminiscent of well known behaviour in a wide range of fields where spatial extent is important, for instance turbulence, pattern formation (such as dislocation dynamics, intermittency in space-time, spatial patterns), Josephson junction arrays, neural dynamics, coupled systems of optically bistable devices, electron-hole plasmas and even parallel computing and evolutionary biology! [1]

So *complex systems research* has emerged as a field of vital importance in the last couple of decades, especially with the advent of high powered computing. It has had impact on several disciplines, as it investigates fundamental dynamic processes which occur in areas of physics, chemistry, and biology. This field of research involves the identification of dynamic processes and concepts which can be employed to understand large families of complex systems. It searches for principles of synthesis of dynamic processes at one level, which produce important dynamic phenomena at another collective level. Thus, in general terms, this field of research seeks to describe and understand the origins of complex dynamic activities and to uncover the methods of modelling and possibly controlling various types of hierarchical dynamics.

One important prototype of extended complex systems is a nonlinear dynamical system with spatially distributed degrees of freedom, or alternately a spatial system composed of large numbers of low dimensional nonlinear systems. There are two principal variations of this scheme: first, the local dynamics can be simple and the interactions complicated in form; secondly, we can have complex local dynamics and simple mode of transmitting information between 'sites' (or 'degrees of freedom'). Here we will be concentrating on the latter situation where locally there is strong nonlinearity. The basic ingredients of such systems are: (i) creation of local chaos or local instability by a low dimensional mechanism and (ii) spatial transmission of energy and information by diffusion or transport (this interconnectivity can work as a global organizing principle in composite systems).

In principle spatio-temporal physics is completely described by the appropriate partial differential equations (PDE). Unfortunately, though, such equations are extremely difficult to solve, and practically intractable without simplifying assumptions and without employing distributed computing techniques. In order to follow complex systems dynamically over long times and over large spatial extents without resorting to heavy duty computing, physicists have used their imagination and evolved a hierarchy of models based on different discretization schemes. These computationally efficient schemes capture some of the essential physics of a very extensive range of natural phenomena. Two of the most succesful schemes are cellular automata (CA) models where space, time and state variables are all discrete, and lattice dynamical systems (LDS) where space and time are discrete and state variables take continuous values. These models have led to the discovery of so many novel phenomena that they are now not merely the poor man's PDE, but bonafide models of complexity in space-time dynamics in their own right.

The aim of all these models is to capture in increasing degree the complexity inherent in nature and provide fitting descriptions of large interactive systems comprising of many elements. In particular, they address two widely occuring phenomena: *spatial fractality*, i.e., the spontaneous evolution of self similarity in space, and *low frequency noise* or $1/f$ noise which represents the lack of a natural scale in time, therefore a kind of "fractality" in time. It is observed that many functional dependences, arising in a variety of experimental contexts, look like straight lines when plotted on a log-log scale, over a significantly large range. So these functions do not have natural length/time scales, that is, they are scale invariant or self similar. This is exemplified in space by spatial fractality, which is ubiquitous in nature, present as it is in spatial structures as diverse as mountain ranges, coastlines, clouds, colloidal aggregates, patterns of fracture, dielectric breakdown, porosity of soil and branching of roots [2]. In the temporal regime, we have $1/f$ noise: $S(f) \sim 1/f$, where f is the frequency and $S(f)$ is the power. (This is also known as "flicker noise" in astronomy, in the context of light from quasars for instance [3]). Such "noise" (or from some viewpoint "signal") is characterized by variations on all timescales, and appears in events as varied as resistance fluctuations, sand flow in an hourglass and even traffic and stock market movements [3]. The generality and ubiquity of these two phenomena have puzzled scientists for a long time. The study of fractality, though, is mostly confined to the level of characterization (for example plotting results of experiments on log-log scales and associating the slope with some kind of fractal dimension). What theorists would really like to do of course, is to find the *dynamical origin* of these phenomena. So, the success of many of the models suggested should be judged with respect to their ability to *dynamically yield space-time fractality.*

In this article we will focus on a class of complex networks, proposed very recently by Sinha and Biswas [4]. These models incorporate adaptive dynamics on a lattice of nonlinear (typically chaotic) elements, and are relevant as prototypes for complex nonlinear self regulatory processes in physics and biology. We will describe the wealth of spatiotemporal structures these models yield, and characterize their "phases" and scaling patterns.

2. Model

Here we describe the essentials of the proposed class of models, which falls under the general category of lattice dynamical systems. First, we focus on the *one dimensional unidirectional* model where time is discrete, labelled by n, space is discrete, labelled by i, $i = 1, N$, where N is system size, and the state variable $x_n(i)$ (which in physical systems could be quantities like energy, velocity, pressure or concentration) is continuous. Each individual site in the lattice evolves under a suitable nonlinear map $f(x)$. For instance, the local map $f(x)$ can be chosen to be the logistic map, which has widespread relevance as a prototype of low-dimensional chaos, i.e.

$$f(x) = 1 - ax^2,$$

$x \in [-1.0,\ 1.0]$, with the nonlinearity parameter a chosen in the chaotic regime, for instance $a = 2.0$ in all numerical experiments in [4]. Another possible choice in the local dynamics is the circle map which is relevant in systems involving nonlinear oscillatory behaviour [6, 7]. This is given by

$$f(x) = x - \frac{\kappa}{2\pi}\sin(2\pi x)$$

with κ determining the strength of the nonlinearity.

Now, on this nonlinear lattice a self regulatory threshold dynamics is incorporated. The adaptive mechanism is triggered when a site in the lattice exceeds the critical value x_c i.e., when a certain site $x_n(i) > x_c$. The supercritical site then relaxes (or "topples") by transporting its excess $\delta x = (x_n(i) - x_c)$ to its neighbour as follows:

$$\begin{aligned} x_n(i) &\to x_c \\ x_n(i+1) &\to x_n(i+1) + \delta x . \end{aligned} \tag{1}$$

This algorithm thus induces a unidirectional nonlinear transport down the array (by initiating a domino effect). The boundary is open so that the "excess" may be transported out of the system. Note that the adaptive ("toppling") mechanism in the model is locally conservative, whereas the intrinsic dynamics of the elements is dissipative. This kind of threshold mechanism imposed on local chaos, makes the above scenario especially relevant for certain mechanical systems like chains of nonlinear springs, as also for some biological systems, such as synapses of nerve tissue (note that individual neurons display complex chaotic behaviour and have step function-like reponses to stimuli).

algorithm [8], or the "sandpile" model, which gives rise to self organized criticality (SOC). This model is however significantly different, the most important difference being that the self regulatory mechanism now occurs on a nonlinear "substrate", i.e. there is an "intrinsic" or "internal" deterministic dynamics at each site. Further, the state variable analogous to the integer "height" variable z in the sandpile model is continuous here. All this accounts for enhanced complexity and this system is thus capable of exhibiting a wider repertoire of dynamics. So, unlike 1D-SOC, the one dimensional model here is not trivial and can give rise to many interesting features, including a host of dynamical phases.

The dynamics depends on the algorithm for autonomously updating each site and propagating threshold coupling between sites. The first possible case is one where these two evolutionary steps are carried out separately. Here the system is allowed to relax completely before each chaotic update. One self-regulatory update of all existing supercritical sites in the lattice, performed according to (1) constitutes one relaxation time step. This relaxation step may in turn create some other supercritical sites. After Δ such relaxation steps, the system undergoes the next chaotic update. In some sense then, time n associated with the chaotic dynamics is measured in units of Δ. When Δ is large enough $(\Delta \gtrsim O(N))$ as in [4], the adpative dynamics begins after each step in the site dynamics and continues till the system has reached a steady state where all sites are less than critical, i.e. all $x(i) \leq x_c$, and the system is stationary, after which the next step in the site dynamics takes place. So the time scales of the two dynamics, the intrinsic chaotic dynamics of each lattice site and the adaptive relaxation, are adiabatically separable. The relaxation mechanism is much faster than the chaotic evolution, and this enables the system to relax completely before the next chaotic iteration. The complete dynamical picture then is as follows: after every time step in the chaotic evolution, n, the system undergoes a self regulatory relaxation leading to stable "under-critical" configurations $\{x_n(i)\}$, after which the next chaotic iteration of the lattice takes place, governed by the nonlinear evolution mapping: $x_{n+1}(i) = f(x_n(i))$. One can also consider the local chaos in the maps to be an "intrinsic" or "internal" perturbation (as opposed to "external" perturbation in the "sandpile" model, which is effected by dropping "sand" from outside). In this sense then, the system above (as its allowed to relax fully between chaotic updates) operates in the "dilute" perturbation limit.

A random driving force can also be introduced in the model. Under this, the system is perturbed at some site j in the lattice: $x_n(j) \rightarrow x_n(j) + \sigma$ where σ is the strength of the perturbation and j is chosen at random. The spatial extent of the perturbation can be gauged by the number of such random sites affected at any given update (which we denote by N_r). The random driving force is operative at time scales comparable to the chaotic dynamics. When this is much slower than the adaptive dynamics, as in [4], the scenario is similar to the SOC algorithm, where the driving force (perturbation) is very dilute, and the system is allowed to relax completely before the next pertubation and time is usually measured in units of the perturbing force (for example, in units of grains of "sand" added in the sandpile model of SOC) and the configurations studied are the relaxed configurations after each perturbation step. This is quite different from the coupled map lattice (CML) dynamics [9], where the coupling is incorporated in the map evolution step.

At the other end of the spectrum is the limit of very small Δ—which is practically analogous to CML. Here the local nonlinear dynamics and chaotic updates take place almost simultaneously. It is evident that lowering Δ essentially allows us to move from a picture of adiabatically separated adaptive processes to one where adaptive processes overlap, and disturbances do not die before the subsequent chaotic update, i.e. *the tails of the relaxation processes carry over to the next event and there are "interference" effects.* The effect of relaxation time on dynamical characteristics was investigated in detail in [7].

The relevant parameters in the model are the local nonlinearity parameter, (for instance, a in the case of logistic maps, and κ in the case of circle maps), strength of perturbation, σ and the relaxation time Δ. The simulations are done with random initial conditions for the $x(i)$ and all transients are allowed to die. Two of the most interesting features to be examined are (i) the temporal evolution of local quantities, such as the individual sites $x_n(i)$, and (ii) the temporal evolution of globally defined quantities, such as "avalanches", which are defined as the total number of "active" sites, i.e. sites that have "toppled" during the adaptive relaxation, denoted by s. The spatial aspects of interest are the distribution of $x(i)$ and the presence of clustering and coherence in space as indicated by the cluster distribution at any point in time (or snapshot). Besides these generic spatio-temporal measures, some other quantities may be of interest in certain contexts. For instance, the dynamics and distribution of the "drop rate", i.e. the "excess" transported out of the lattice edge during relaxation, may be significant in some physical situations (such as in phenomena reminiscent of the "dripping faucet" scenario).

Note that x_c is the most significant parameter in the system. For instance, for the case of the local logistic map ($f(x) = 1 - 2x^2$), where $-1.0 < x_c < 1.0$, numerical simulations show the presence of the many "phases" in x_c space [4]. For example, for $x_c \leq 0.5$ the dynamics goes to a fixed point. When $0.5 < x_c < 1.0$, the dynamics is attracted to *cycles* whose periodicities depend on the value of x_c. By tuning x_c one thus obtains cycles of varying orders. So in a sense, this model (by variation of a single parameter) provides a wide repertoire of dynamics for lattice dynamical systems, just as the logistic map does for one dimensional systems (by variation of the nonlinearity parameter). Thus it may be used as a paradigm for complex behaviour in extended systems.

Note another important feature here. If the same adaptive dynamics was imposed on a random lattice, that is a bunch of noisy elements, we would not have recovered any of the above mentioned phases. For these phases to occur we crucially require a *deterministic dynamics*, even if entirely chaotic. *The difference between stochastic noise and deterministic chaos is of vital importance here.*

We can also introduce the bidirectional model. This is given by the following algorithm: if $x_n(i) > x_c$, the supercritical site then relaxes to x_c by transporting the excess $(x_n(i) - x_c)$ equally to its two neighbours:

$$\begin{aligned} x_n(i) &\rightarrow x_c \\ x_n(i+1) &\rightarrow x_n(i+1) + \delta x \\ x_n(i-1) &\rightarrow x_n(i-1) + \delta x \end{aligned} \tag{2}$$

where $\delta x = (x_n(i) - x_c)/2$. Now disturbances in the course of self-regulation, can spread like "ripples" and "refract". The dynamics of the bidirectional model looks, by and large, like a "fuzzier" or "noisier" version of its unidirectional counterpart. The details will be discussed in the following sections.

Before we move on to a specific example in § 3, we will write down the generalized model in d dimensions with m state variables. Let each state variable defined at a certain site in the lattice be denoted as $x^j(\mathbf{i})$, where $\mathbf{i}$ is a d dimensional vector representing the coordinates of the site in the d dimensional lattice, and j represents a particular state variable, taking values from 1 to m. If a certain state variable, say $x^j(\mathbf{i})$ at a certain site $\mathbf{i}$ in the d lattice exceeds the critical value, x^j_{c}, then the excess is transported equally to its neighbours. (Note that in general we may have different critical values for different state variables, and these are given by m numbers, $x^j_{\mathrm{c}}, j = 1, 2, \ldots, m$). Also note that in d dimensions the number of neighbours (assuming isotropic transport) is $2d$. Given these the adaptive dynamics then is

$$\begin{aligned} x^j_n(\mathbf{i}) &\rightarrow x^j_{\mathrm{c}} \\ x^j_n(\mathbf{i}+\boldsymbol{\delta}) &\rightarrow x^j_n(\mathbf{i}+\boldsymbol{\delta}) + \delta x, \end{aligned} \tag{3}$$

where the quantity δx gives the excess, which is equally distributed among the $2d$ neighbours: $\delta x = (x^j(\mathbf{i}) - x^j_{\mathrm{c}})/2d$. The vector $\boldsymbol{\delta}$ is d-dimensional, comprising of only one nonzero element ($= 1$ or -1) and it gives the coordinates of the nearest neighbours of the critical site to which the excess is transported.

Some other variations to the model are also possible. For instance, we can investigate the effect of changing boundary conditions. In particular, we can incorporate the adaptive dynamics on a lattice with periodic boundary conditions, that is on a ring of nonlinear elements. Further, for instance, one can implement the adaptive dynamics (or transport) on a hierarchical lattice, in order to evoke scaling phenomena reminiscent of energy cascade in turbulent media. In general then, several variations of this class of models can be invoked in order to tailor the generalities to suit the specific physics at hand and account for particular phenomena.

In this article we will focus on one representative example, namely, adaptive dynamics on logistic maps. We give a detailed description of the system and its emergent phenomena in the section below.

3. Adaptive dynamics on logistic maps

In this model the local on-site dynamics is given by logistic maps. The specific form of the logistic map used (without any loss of generality) is: $f(x) = 1 - ax^2$, $x \in [-1, 1]$, operating in the chaotic regime, i.e., $a = 2$. We deal with the case of Δ large, which allows complete relaxation of the active sites before the subsequent chaotic update. First, we describe the unidirectional case, with and without external noise, and then describe the bidirectional model.

3.1 *Unidirectional model: Phenomenology*

As noted in § 2, the most significant parameter in the system is the critical x_{c} ($x_{\mathrm{c}} \in (-1, 1)$), and by tuning x_{c} one obtains the following phases: The first phase is the fixed point region which occurs when $x_c < 0.5$. Here the system goes to a coherent state where all sites $x_n(i) = x_c$ for all times (after a short transience). This spatiotemporal invariance is achieved for all perturbation strengths and system sizes. The size of the avalanches is always equal to $N(N+1)/2$ and the duration of the avalanches equal to N [4]. In this parameter regime, then, the adaptive mechanism suppresses the underlying chaos in the lattice and yields spatiotemporal regularity.

The model (in the range $x_c < 0.5$) may then be considered as a tool for "controlling" [10] an ensemble of chaotic elements, as it can very effectively force the system to a temporally invariant and spatially coherent state: $x_n(i) = x_c$, $i = 1, \ldots, N$ for all n. Note that this spatiotemporal control is robust with respect to system size and perturbation strengths. Couched in the language of control theory, we can consider x_c to be the "desired state" of the system [10]. The "error signal" is then determined by the difference between the existing state and the desired state. A positive error signal triggers off the self regulatory feedback mechanism which drives the system to adaptively control its dynamics back to the desired state (when the error is zero).

When $x_c = 0.5$ we still have a coherent state with all $x(i) = x_c$. When there is no noise the avalanches are of size zero as there are never any active sites in the lattice, as $x = 0.5$ is a fixed point of the map $f(x)$. Under a random driving force ($N_r \neq 0$ and $\sigma \neq 0$) the

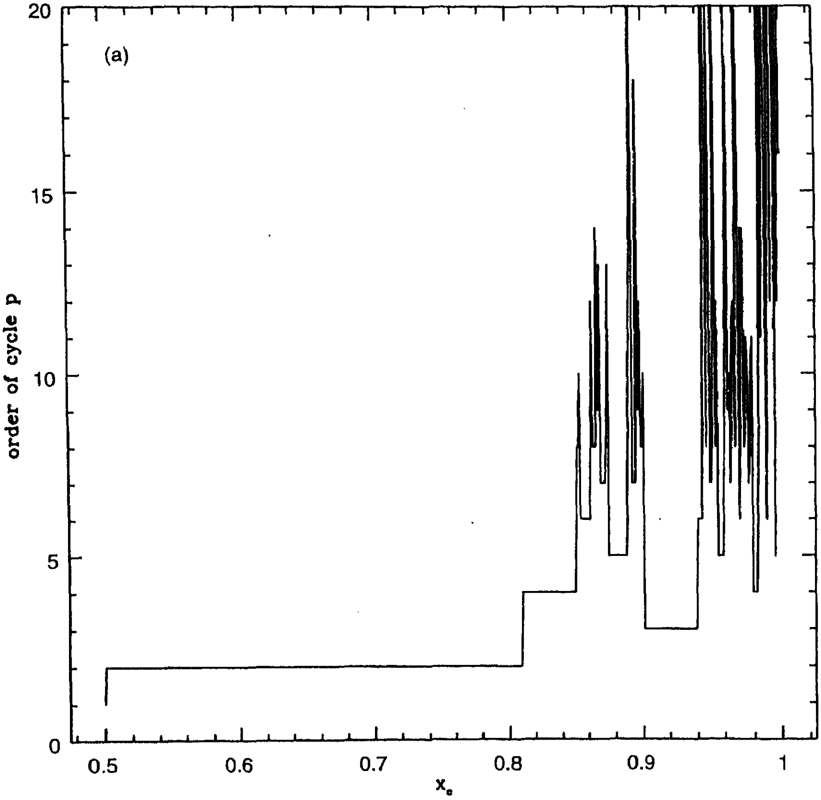


Figure 1a. *(Continued)*

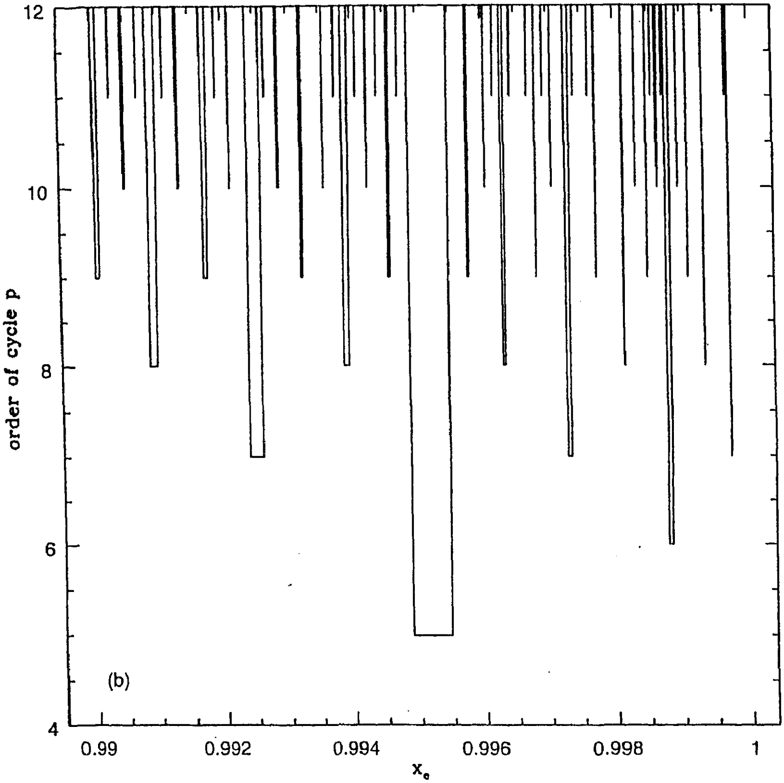


Figure 1. Order p of the cycle obtained at various values of the parameter x_c, in a unidirectional adaptive model on a lattice of logistic maps ($\sigma = 0$), depicted for the ranges: (**a**) $x_c \in [0.5, 1.0]$ (k up to 20 is shown) and (**b**) $x_c \in [0.99, 0.9999]$ (k up to 12 is shown).

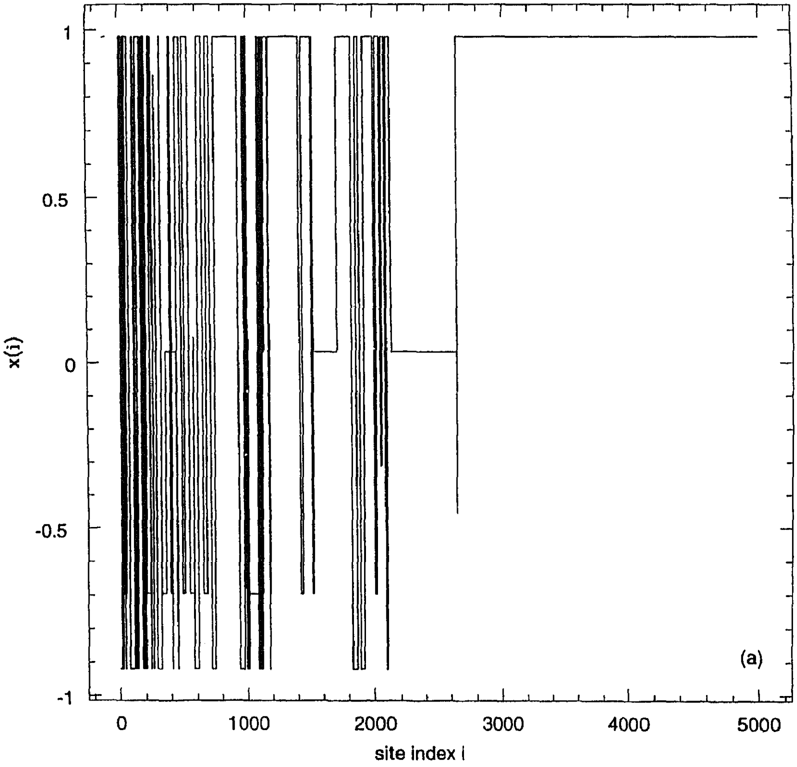
avalanches are triggered off, and they have a random distribution of sizes, as the avalanche size is simply determined by the distances of the perturbed sites from the edge of the lattice, and from each other, and these are statistically random quantities.

When $0.5 < x_c < 1.0$, the temporal evolution of the lattice sites (and avalanches) is attracted to a *cycle* whose periodicity (order) p depends on x_c [4]. For example, by tuning x_c one obtains the following dynamical phases: for $0.5 < x_c \leq 0.809..$, we get 2-cycles, for $0.809.. < x_c < 0.85$ we get 4-cycles, for $x_c \sim 0.86$ we get 6-cycles, for $x_c \sim 0.88$ we get 7-cycles, for $x_c \sim 0.9$ we get 10-cycles and for $x_c \sim 0.98$ we get 4-cycles, and so forth. When there is no external perturbing force ($\sigma = 0$) these cycles are *exact*. From figure 1 (which gives the order of the cycle, p, supported at various values of x_c) it is clearly evident that the system yields a rich repertoire of cyclic patterns. Note here that it is possible to obtain a detailed and rigorous analytical picture of the spatial and temporal characteristics of the various cycles: see [5]. This analysis [5] provides a dynamical basis for the various phases found in this model. Further, the analysis also gives exact scaling relations for the position and width of periodic windows in x_c space. Due to constraints of space, however, we will not touch upon those analytical results in this article, but concentrate instead on the numerical results found principally in [4].

Temporally, then, we obtain cycles of varying orders. Now, spatially, we would like to discern the degree of "coherence" in the spatial profile at any instant of time (or snapshot) by examining the spatial patterns in $x_n(i), i = 1, 2, \ldots, N$. For instance, we can quantify spatial clustering in the landscape by defining the size c of a spatial cluster as the number of consecutive sites for which the inequality $|x_n(i+1) - x_n(i)| < \epsilon$ holds, where ϵ is a small number. (The results are not very sensitive on the choice of ϵ). Very large cluster sizes, i.e. $c \sim O(N)$, indicate a smooth, almost coherent, profile, whereas profiles with maximum cluster size $c_{\max} \sim 10$ reflects a rough profile.

It is observed that for values of x_c supporting low order periodicities, there often exists one very large cluster. In addition to this large (isolated) cluster, there exist many smaller clusters which are distributed approximately as a power law. For instance, for the case of $x_c = 0.98$ which supports a 4-cycle, we usually find one large cluster, $c_{\max} \sim N/2$, and a host of smaller clusters (see figure 2). The coarse grained distribution of the small cluster sizes, $P(c)$, appears to be a power law: $P(c) \sim c^{-\phi_c}$, with exponent ϕ_c depending on the specifics of the model [11]. On addition of noise the spatial profile gets more uneven, as the clusters break into smaller clusters. For instance, for $x_c = 0.98$, $\sigma = 0.2$, $c_{\max} \sim N/4$, that is the largest cluster is around half as large as in the $\sigma = 0$ case.

For larger x_c supporting cycles of high order (for example, $x_c = 0.99$) we find that the landscape consists of several small clusters of size c, occuring over a range of scales,



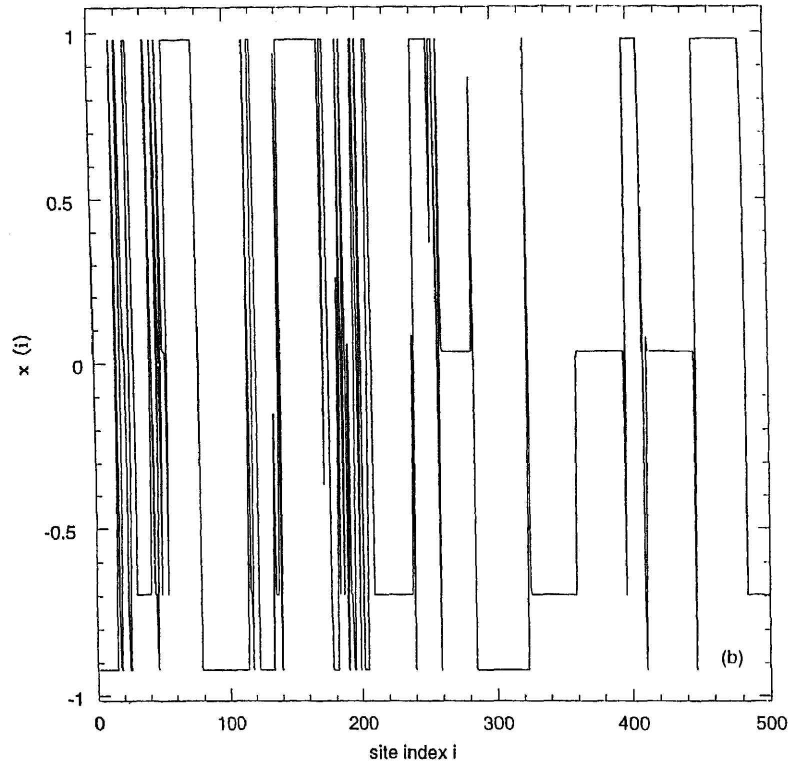


Figure 2. A plot of state variable $x(i)$ vs site index i, in a unidirectional adaptive model on a lattice of logistic maps, at $x_c = 0.98$, $\sigma = 0$, for a lattice of size $N = 5000$. In (**a**) we depict the entire lattice: $i = 1, \ldots, 5000$. Note the existence of one very large cluster (all sites with $2648 \leq i \leq 5000$ have $x(i) = 0.98$). In (**b**) we depict a blown up version of the section of the lattice with $1 \leq i \leq 500$. Note the range of cluster sizes.

up to $c_{\max} \sim 10$. The profile is now very rough (see figure 3). Also, note that the addition of noise now does not appreciably alter the picture. The distribution $P(c)$ follows a power law: $P(c) \sim c^{-\phi_c}$, where the exponent ϕ_c depends on the particulars of the model.

3.2 *Unidirectional dynamics under weak noise*

Very interesting phenomena emerge on the addition of weak noise. When noise is very weak (i.e., σ is small, and number of affected sites, N_r is small, say $N_r/N < 0.1$) then the dynamics remains regular, especially for low x_c values supporting low order cycles. On adding moderate noise, $\sigma \sim 0.01$, and $N_r/N \sim 0.1$, we get interesting results for certain x_c values: namely large x_c with a reasonably wide window of low periodicity associated with it. At these parameters the system yields a well defined $1/f$ spectrum: $S(f) \sim 1/f^{\phi}$, $\phi \sim 1.0$. So the time profile now has variations on all scales, with interesting "cyclic" looking broad trends. Figure 4 displays a representative example of $x_c = 0.98$, $\sigma = 0.01$, $N_r = 1$ and $N = 20$ which yields a power spectrum with $\phi \sim 0.93 \pm 0.06$. Also notice

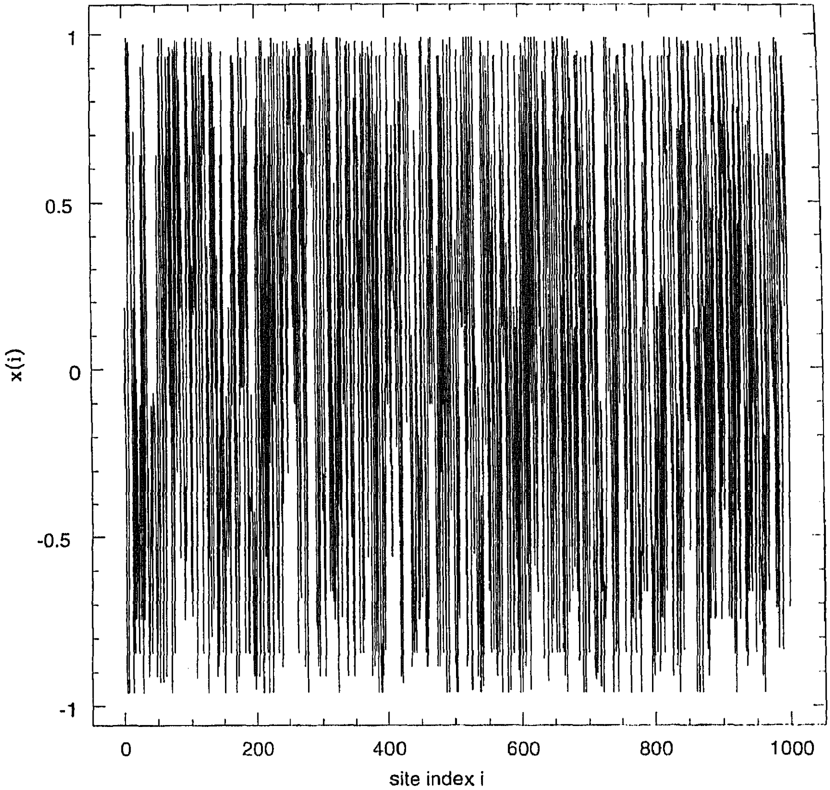


Figure 3. A plot of state variable $x(i)$ vs site index i, in a unidirectional adaptive model on a lattice of logistic maps, at $x_\mathrm{c} = 0.99$, $\sigma = 0$, for a lattice of size $N = 1000$. Note the rough looking profile.

that there is a large peak in the spectrum at $f = \frac{1}{4}$, which is the period associated with x_c when $\sigma \sim 0$.

This kind of temporal scale invariance is reflected in the distribution of avalanche size s, which falls as a power law: $P(s) \sim s^{-\phi_s}$. It is also observed that $P(s)$ is considerably peaked at large values of s ($s \sim O(N)$). This hump at large s reflects the very large events, which occur approximately periodically, with the period given by the $\sigma = 0$ model. Figure 5 gives the $P(s)$ vs s curve for the representative case of $x_\mathrm{c} = 0.9$. The exact value of the exponent ϕ_s depends on the system parameters, and usually has values close to 1.0 (here for $x_\mathrm{c} = 0.9$, $\phi_s \sim 1.2 \pm 0.2$).

A further quantity of interest would be the sizes of active clusters, k, at any given instant of time (or snapshot) after the relaxation has taken place. This is determined by the number of consecutive sites that move (or become active) during the course of an adaptive process. This number can vary from 1 to N.

Numerical experiments indicate that the distribution $P(k)$ follows a power law: $P(k) \sim k^{\phi_k}$. The exponent ϕ_k depends on the particulars of the system and varies in the range $\sim$ 1.0 to 2.0. For instance, from figure 6 (which shows a representative example of $x_\mathrm{c} = 0.88$, with $P(k)$ obtained by averaging over 2×10^4 realizations) it is clearly evident

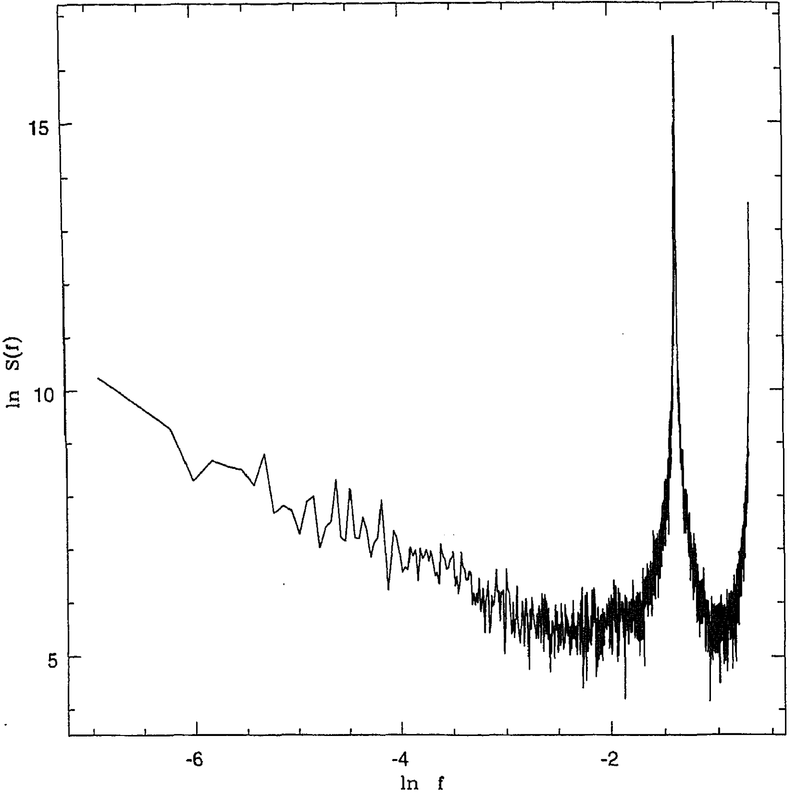


Figure 4. Power spectrum of avalanches in a unidirectional adaptive model on a lattice of logistic maps, with $x_c = 0.98$, $\sigma = 0.01$, and $N_r = 1$, $N = 20$. We have averaged over 8 time runs of 2048 each. The abcissa has $\ln f$, where f is the frequency ($f \in (0, 0.5]$), and the ordinate has $\ln S(f)$, where $S(f)$ is the power.

that there exists well defined scaling, with the exponent $\phi_k \sim 1.1$. This scaling reflects the fact that responses to perturbation (or energy dissipation) in this model takes place over a range of length scales, starting from small local rearrangements to large events where the transport activity is at a wider scale involving all the elements of the lattice.

For large values of x_c supporting very high order cycles, the spectra is noisy and flat ($S(f) \sim 1/f^0$) even for very weak noise. Also, the distribution of avalanche size s is a gaussian. So the system now behaves more like a bunch of random elements.

3.3 *Bidirectional model: Phenomenology*

Bidirectional transport mimics its unidirectional counterpart closely, but seems to be a noisier or "fuzzier" version of it. Thus we find that for $x_c \leq 0.5$ all the sites are randomly distributed with a small deviation around the mean which is almost equal to x_c. The avalanches are all of the maximum possible size $s_{\max} = \Delta \times N$. When $0.5 < x_c \lesssim 0.8$ the sites get attracted to noisy 2-cycles, as is evident through the power spectrum of the

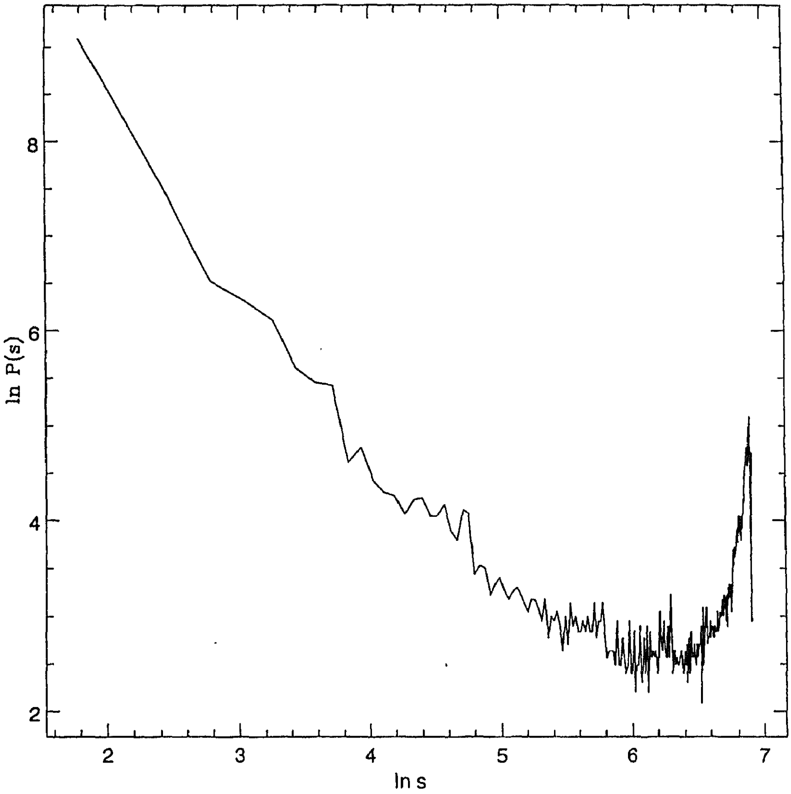


Figure 5. Plot of the distribution of avalanche sizes: $\ln P(s)$ vs $\ln s$, in a unidirectional adaptive model on a lattice of logistic maps, for $N = 1000$, $x_c = 0.9$, and $\sigma = 0.1$. We sample 2×10^4 configurations.

temporal evolution of the individual sites which exhibit a large (but not clean) peak at 1/2. The avalanches evolve as exact 2 cycles, alternating between 0 and $s_{\max}$. Further, as we tune x_c we obtain a noisy 4-cycle, and so forth.

As x_c gets larger we have a rather complex scenario with a host of scaling properties emerging in space and time. Firstly, for certain x_c values, namely large x_c with a reasonably wide window of low periodicity associated with it, we observe a time profile having variations over a wide range of scales, giving rise to a well defined $1/f$ spectrum. For instance, in the example shown in figure 7, the system ($x_c = 0.98$) displays a prominent low frequency spectrum, with $\phi \sim 1.54 \pm 0.06$. Note also, the existence of a sharp peak at $f = \frac{1}{4}$, which is the frequency of the exact cycle associated with the unidirectional $\sigma = 0$ system. (It is then evident from comparing figures 4 and 7, that the spectrum for the bidirectional model is strikingly similar to that obtained from unidirectional dynamics under weak noise).

The active clusters, typically, now come in all sizes, indicating again that activity is taking place on several scales, starting from small rearrangements to large events sweeping through the lattice. The distribution of active cluster sizes $P(k)$, displays a clearly defined power law: $P(k) \sim k^{-\phi_k}$. The exponent depends on the particular system

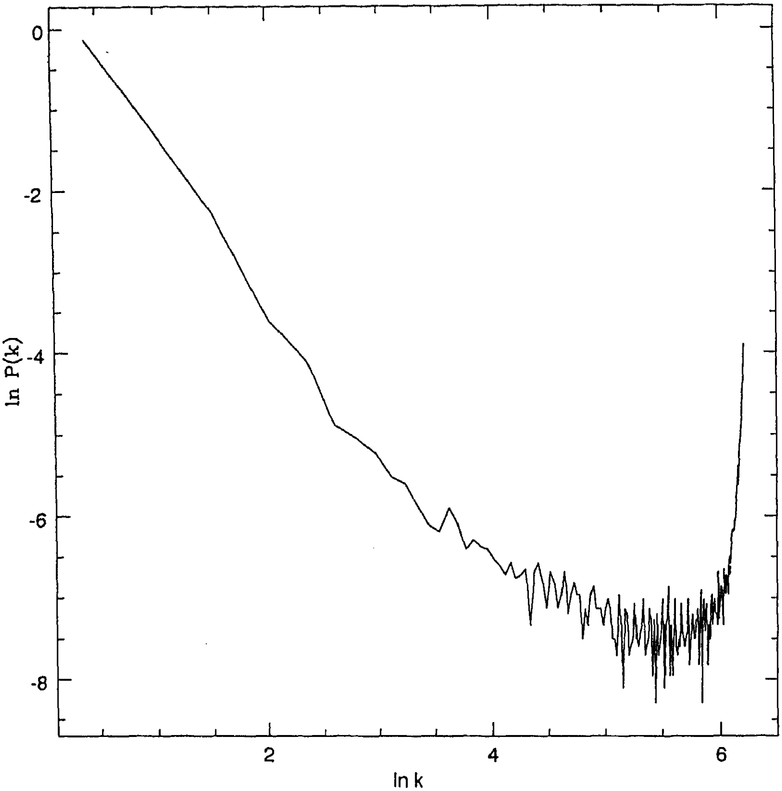


Figure 6. Plot of the distribution of active clusters: $\ln P(k)$ vs $\ln k$, in a unidirectional adaptive model on a lattice of logistic maps, for $x_c = 0.88$, $\sigma = 0.2$, $N = 500$. We have averaged over 2×10^4 configurations. Notice the enhanced probability of finding large active clusters, as evident from the rising $P(k)$ around $k_{\max} \sim N/5$.

we examine. For instance, for $x_c = 0.9$ and σ in the range 0.0 to 0.2, we obtain $\phi_k \sim 4.0$, and for $x_c = 0.98$, for noise levels $\sigma = 0.0$ to 0.2, we obtain $\phi_k \sim 2.0$. This scaling reflects the fact that responses to perturbation (or energy dissipation) in this model takes place over a range of length scales.

A further quantity of interest would again be the distribution of avalanches. Here we focus only on the coarse grained distribution, as the range of values of s, from 0 to $s_{\max} = \Delta \times N$, is rather large. First, it is observed that there is a considerable hump at large values of s ($s \to s_{\max}$). This is related to the fact that there are large events which occur approximately periodically (with the same period as that of the system's unidirectional counterpart). For instance, for $x_c = 0.98$, approximately one in every four avalanches is very large. (Such avalanches actually take infinitely long to settle down, and so their size is bound only by finite Δ, i.e. $s = s_{\max} = \Delta \times N$.) For small sizes, it was found that the distribution falls as a power law: $P(s) \sim s^{-\phi_s}$. While this power-law trend persists for weak noise, under addition of moderate noise the system often yields an exponentially falling distribution: $P(s) \sim e^{-\phi_s s}$. Further, very high x_c ($x_c \to 1.0$) also yields such exponentially decaying $P(s)$.

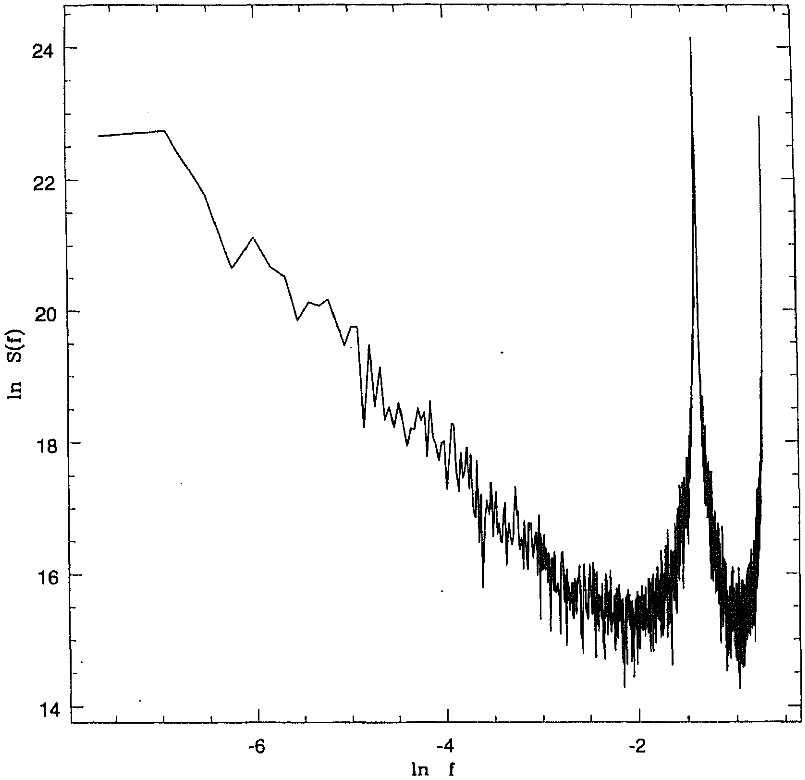


Figure 7. Power spectrum of avalanches in a bidirectional adaptive model on a lattice of logistic maps, for $x_c = 0.98$, $N = 20$, and $\sigma = 0.01$, $N_r = 3$. We average over 8 time runs of 2048 each. The abcissa has $\ln f$, where f is frequency ($f \in (0, 0.5]$), and the ordinate has $\ln S(f)$, where $S(f)$ is the power.

Spatially, we obtain power law-like distributions for spatial cluster sizes c. We find a host of small clusters up to a size of $c_{\max} \sim N/25$. Note that the addition of noise does not appreciably alter the roughness of the profile. Note, also, that unlike its unidirectional counterpart, there are no large clusters here. The coarse grained distribution of the small cluster sizes $P(c)$ appears to be a power law, with exponent ϕ_c depending on the specifics of the model.

4. Summary

In this article, we have described a novel class of models incorporating adaptive dynamics on a lattice of nonlinear (typically chaotic) elements. We demonstrate the wide repertoire of dynamical patterns these systems are capable of supporting. The spectrum of emergent spatio-temporal phenomena ranges from spatio-temporal fixed points and cycles, to parameter regimes marked by scaling properties, as manifest in clearly defined $1/f$ spectral characteristics and power law distributions of spatial quantities. The capacity of these models to yield wide ranging dynamical patterns by tuning just one parameter, underscores their utility as versatile paradigms of spatio-temporal complexity in extended systems.

References

[1] J Crutchfield and K Kaneko, in *Directions in chaos* edited by Hao Bai-Lin (World Scientific, Singapore, 1987) and references therein

[2] B B Mandelbrot, *Fractal geometry of nature* (Freeman, San Francisco, 1982)
T Vicsek, *Fractal growth phenomena* (World Scientific, Singapore, 1989)

[3] W H Press, *Comm. Mod. Phys.* **C7**, 103 (1978)

[4] S Sinha and D Biswas, *Phys. Rev. Lett.* **71**, 2010 (1993)
S Sinha, *Int. J. Mod. Phys.* **B9**, 875 (1995)

[5] S Sinha, *Phys. Rev.* **E49**, 4832 (1994)

[6] K Kaneko, *Physica* **D54**, 5 (1991) and references therein

[7] S Sinha, *Phys. Lett.* **A199**, 365 (1995)

[8] P Bak, C Tang and K Weisenfeld, *Phys. Rev. Lett.* **59**, 381 (1987); *Phys. Rev.* **A38**, 364 (1988)

[9] R Kapral, *J. Math. Chem.* **6**, 113 (1991) and references therein

[10] B Huberman and E Lumer, *IEEE Trans. Circuits and Systems* **37**, 547 (1990)
S Sinha, R Ramaswamy and J Subba Rao, *Physica* **D43**, 118 (1990) and references therein

[11] Note that the probabilities reported throughout ($P(c)$, $P(s)$ and $P(k)$) are in arbitrary units, and are not normalized to unity. Of course, this does not affect any of the conclusions reached by inspecting the graphs

Control of spatiotemporal chaos: A study with an autocatalytic reaction-diffusion system

NITA PAREKH, V RAVI KUMAR and B D KULKARNI
Chemical Engineering Division, National Chemical Laboratory, Pune 411 008, India

Abstract. The characterization of chaotic spatiotemporal dynamics has been studied for a representative nonlinear autocatalytic reaction mechanism coupled with diffusion. This has been carried out by an analysis of the Lyapunov spectrum in spatially *localised* regions. The linear scaling relationships observed in the invariant measures as a function of the *sub-system* size have been utilized to assess the controllability, stability and synchronization properties of the chaotic dynamics. The dynamical synchronization properties of this high-dimensional system has been analyzed using suitable Lyapunov functionals. The possibility of controlling spatiotemporal chaos for relevant objectives using available noisy scalar time-series data with simultaneous self-adaptation of the control parameter(s) has also been discussed.

Keywords. Spatiotemporal chaos; reaction-diffusion systems; autocatalysis; Lyapunov exponents; synchronization and control; parametric self-adaptation.

PACS Nos 05.70; 82.20; 47.54; 05.45

1. Introduction

Spatially extended nonlinear systems are known to exhibit a wide variety of complex dynamics ranging from stable stationary states to coherent travelling waves, solitons and spatiotemporal chaos. Studies on this topic have been reviewed by Cross and Hohenberg [1] from a global perspective and their importance in many physical, chemical, biological and social systems have been summarized. To cite a few examples, complex patterns have been experimentally observed in hydrodynamic fluids, nonlinear optics, crystallization of solids, chemically reacting systems, morphogenesis, self-replication of living cells, neural systems, population dynamics, and economics. Even though these systems are high-dimensional due to mechanisms such as diffusion, convection, etc., concurrently occurring in the spatial domain, a wide range of complex phenomena exhibited by these systems may be modelled by simple partial differential equations or coupled maps. Considerable application potential exists if it were possible to devise strategies that would regulate and control these dynamical patterns to exhibit the desired behaviour. This is especially true for systems exhibiting spatiotemporal chaotic patterns which are highly sensitive to the initial conditions and parameter settings [2] and their robust control needs highly efficient and newer techniques of analysis to be developed.

Systems exhibiting chaotic behaviour offer the advantage of controlling the system dynamics in a variety of dynamical patterns with low energy inputs [3]. Typically, an experimenter may wish to stabilize the spatiotemporally chaotic dynamics, or switch it over to a different dynamical pattern (which may also be chaotic), using limited available information in the form of scalar time-series data in the spatial domain. Possible control strategies in such situations would require controllers to be dispersed in the spatial domain with each one perturbing the value of the chosen control parameter(s), or alternatively, introducing appropriate forcing in the form of dynamical waveforms from outside the system. Recent studies using these methodologies [4–7] have shown that it may be possible to suitably tune the spatiotemporally chaotic dynamics to a stabilized mode or a periodic orbit of the homogeneous system, devoid of spatial effects. A dense distribution of the controllers may be necessary to account for the large number of unstable degrees of freedom in high-dimensional systems. However, if some information about the basic mechanism governing the system dynamics is known, then ingenious strategies involving simple localized control algorithms may be employed. For instance, the spatiotemporal patterns observed in cardiac tissues [8], reaction-diffusion models for catalysis [9, 10], and Ginzburg-Landau equation [11–13] arise because of unstable topological defects nucleating spontaneously from an initial disordered state. In such situations it has been shown that the stabilization of even one wave-generating defect may interact and control all the others right up to the system boundary [13]. In the event when the spatiotemporal system is subjected to noise and a periodic forcing function, the cooperative effects that may arise between the inherent nonlinear dynamics, the external noise and periodic forces may stabilize the system dynamics due to the phenomenon of stochastic resonance. In fact, in a recent experimental study with linearly coupled array of nonlinear oscillators, it has been shown that ordered periodic states can be achieved by the optimization of noise levels and/or forcing frequency [14, 15]. Thus, while working with chaotic systems, even unconventional methods may be rationally applicable for regulating the system dynamics.

The control of nonlinear chaotic systems may also be looked at from a different viewpoint. Given a process from which we monitor scalar time-series data, can we use it to drive a response system so that its dynamics in all the other non-monitored variables becomes synchronized with that of the process? Despite the fact that the chaotic behaviour exhibits rapid decorrelation of nearby orbits, it has been observed that it may be possible to synchronize the response system dynamics to that of the process [16]. Considerable attention has been focussed on this topic in the recent years and conditions for obtaining synchronization have been studied when the response system is: 1) a replica of the process with incorrect knowledge of the initial conditions and 2) a non-replica of the process with unknown initial conditions and incorrect parameter settings. In particular it has been shown that for synchronization in initial conditions to occur, the response system, excluding the driving variables should possess only negative Lyapunov exponents [17–26]. When the response system is a non-replica of the process, synchronization may be possible by suitable perturbations in the response parameter(s) (depending upon local stability considerations) [27, 28] – much in the same fashion as suggested by OGY for controlling chaos [29, 30].

It may be noted that the above synchronization and control studies refer to systems described by low-dimensional chaotic systems. On the other hand, spatio-temporal chaotic systems may be thought of as a network of chaotic elements with a large number of positive Lyapunov exponents, and hence, for their dynamical synchronization a large number of driving signals may be necessary. Furthermore, it would be computationally difficult to characterize the system dynamics and clearly demarcate the stable and unstable directions associated with the attractor because of the high dimensions involved [7, 31–33, 35, 37–39]. It is then worthwhile to see if simpler and practical ways of analyzing these systems can be devised. A possible strategy may be to first assess the complexity and stability properties of the high-dimensional system in a lower dimensional framework. Our earlier studies on this topic have shown that interesting scaling relationships in the system invariant measures (such as Lyapunov exponents, entropy, etc.) as a function of sub-system size exists for the spatiotemporally chaotic dynamics in nonlinear reaction-diffusion systems [40]. An analysis of these relationships may yield valuable information about the requirements of spatial time-series data for the characterization and control of spatiotemporal systems.

In this paper, we shall discuss the control of spatiotemporally chaotic dynamics in nonlinear reaction-diffusion systems for limitations in the knowledge of the system. The reaction-diffusion system considered here belongs to a fairly general class of chemically reacting systems involving nonlinearity by way of autocatalytic steps [45, 49] and in the following section we shall elaborate on the specifics of the system studied. Subsequently, in § 3 we summarize the possible methodology to characterize the spatiotemporal chaotic behaviour in terms of sub-system invariant measures like Lyapunov exponents, entropy, etc. Section 4 studies the synchronization and control of replica and non-replica response systems with simultaneous estimation of the response parameter(s) and § 5 briefly summarizes the scope of the results.

2. Model description

In recent times considerable interest has been aroused in the study of nonlinear reaction-diffusion systems with the experimental observance of complex spatiotemporal patterns, e.g., striped, hexagonal, travelling waves, and spatiotemporal chaos in a simple gel-reactor uncontaminated by convection effects [41, 44]. Pattern formation in these systems arises as a result of spontaneous symmetry breaking transitions with intrinsic length scales dependent on the diffusion coefficients and the reaction rate constants [42, 43]. An interesting class of spatial patterns observed include spots which have an uncanny resemblance to what is observed in many biological systems in the sense that they replicate, grow and die [46, 47]. These replicating patterns, do not have the high degree of sensitivity associated with the chaotic dynamics and hence their control is less difficult. In fact, in an earlier study with a simple autocatalytic reaction-diffusion model, using controllers dispersed in the spatial domain, we have shown that the control of these patterns in desired dynamical states may be possible [48]. In the present study we consider a similar reaction network with an additional mechanistic step involving a new species D, i.e., a three-step parallel autocatalytic reaction mechanism [49] with competing interactions between the chemical species A, D, and B, viz.,

$$A + 2B \longrightarrow 3B, \quad r_1 = k_1 C_A C_B^2, \tag{I}$$

$$D + 2B \longrightarrow 3B, \quad r_2 = k_2 C_D C_B^2, \tag{II}$$

$$B \longrightarrow C, \quad r_3 = k_3 C_B, \tag{III}$$

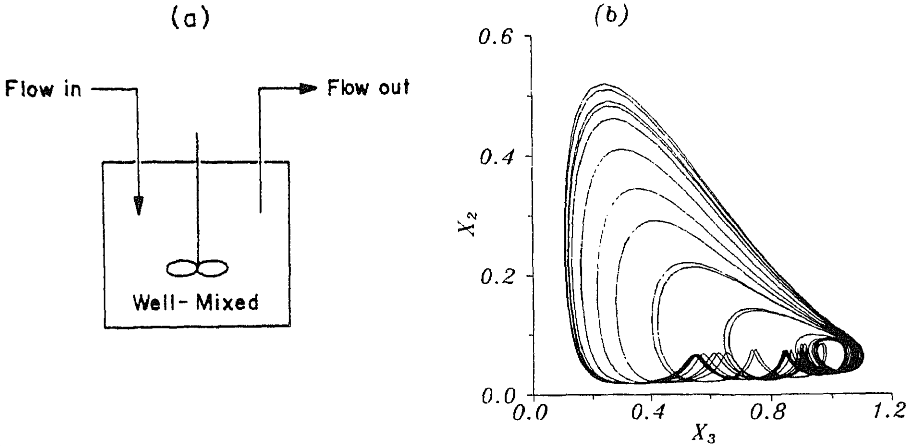


Figure 1. (**a**) Schematic diagram of a well-mixed continuous stirred tank reactor (CSTR). (**b**) Temporal chaos arising for the autocatalytic reaction (1) in a CSTR, shown as a phase-plane plot; $\alpha = 1.5$; $\beta = 2.93$; $Da_1 = 18000$; $Da_2 = 400$; $Da_3 = 80$.

and the rate expressions r_i, $i = 1, 2, 3$ for each step, given alongside. Nonlinear feedback in this mechanism occurs due to the rates of formation of species B being autocatalytic in steps (I) and (II), while, in (III) the effect is inhibitory. For a continuous flow well-mixed cell (shown in figure 1a), the model description for the above reaction takes the following dimensionless form

$$\frac{\mathrm{d}X_1}{\mathrm{d}t} = 1 - X_1 - Da_1 X_1 X_3^2,$$

$$\frac{\mathrm{d}X_2}{\mathrm{d}t} = \beta - X_2 - Da_2 X_2 X_3^2,$$

$$\frac{\mathrm{d}X_3}{\mathrm{d}t} = 1 - (1 + Da_3)X_3 + \alpha(Da_1 X_1 + Da_2 X_2)X_3^2, \tag{1}$$

where X_i, $i = 1, 2, 3$, respectively, represent the dimensionless species concentrations of A, D, and B relative to their concentrations at the inlet to the cell. Here, Da_i's are the dimensionless kinetic parameters associated with the reaction steps (I–III) and the parameters α and β are the feed concentration ratios of species B and D with respect to A. An analysis of the dynamical properties of this cell shows features such as multi-stationarity, oscillatory behaviour and chaos in chosen parameter space of α, β and Da_i's [49]. In figure 1b a typical phase portrait diagram of X_2 vs X_3 depicting the chaotic dynamical behaviour of (1) (with maximum Lyapunov exponent, $\lambda_{\max} \sim 1.36$) is shown. On introducing a diffusive transport mechanism for a system of N coupled autocatalators (1), the system dynamics may be described by the following set of PDE's, viz.,

$$\frac{\partial X_1(\mathbf{r}, t)}{\partial t} = 1 - X_1(\mathbf{r}, t) - Da_1 X_1(\mathbf{r}, t) X_3^2(\mathbf{r}, t) + d_1 \frac{\partial^2 X_1(\mathbf{r}, t)}{\partial \mathbf{r}^2}$$

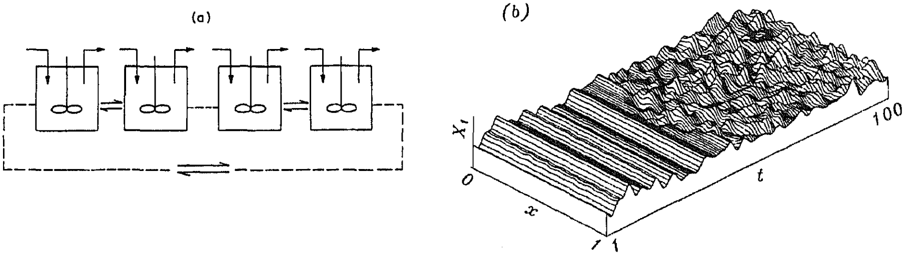


Figure 2. (**a**) Schematic of N CSTRs coupled bi-directionally via diffusion in one spatial dimension with periodic boundary conditions. (**b**) Spatiotemporal chaos arising for a perturbation given at $t = 40$. $D_1 = D_2 = 1.0$; $D_3 = 0.01$; $\alpha = 1.5$; $\beta = 2.93$; $Da_1 = 18000$; $Da_2 = 400$; $Da_3 = 80$; $[X_1$ axis-scale: $(0.0, 0.07)$.

$$\begin{aligned}\frac{\partial X_2(\mathbf{r},t)}{\partial t} &= \beta - X_2(\mathbf{r},t) - Da_2 X_2(\mathbf{r},t) X_3^2(\mathbf{r},t) + d_2 \frac{\partial^2 X_2(\mathbf{r},t)}{\partial \mathbf{r}^2} \\ \frac{\partial X_3(\mathbf{r},t)}{\partial t} &= 1 - (1 + Da_3) X_3(\mathbf{r},t) + \alpha[Da_1 X_1(\mathbf{r},t) + Da_2 X_2(\mathbf{r},t)] X_3^2(\mathbf{r},t) \\ &\quad + d_3 \frac{\partial^2 X_3(\mathbf{r},t)}{\partial \mathbf{r}^2}, \qquad (2)\end{aligned}$$

where $\mathbf{r} \equiv (x, y, z)$ is the spatial vector; and d_i, $i = 1, 2, 3$ are the dimensionless diffusion coefficients of the respective species A, D and B. Autocatalytic reaction-diffusion systems of the type (2) are known to possess a rich variety of dynamical patterns with properties such as excitability, wave reflection and wave splitting, propagating fronts, etc. [50, 51].

In this paper, all studies were carried out for N independent autocatalators coupled via bi-directional diffusion on a one-dimensional spatial lattice (schematically shown in figure 2a). On Euler discretizing the Laplacian, $\partial^2/\partial x^2$, in the spatial dimension x, viz., we obtain

$$\begin{aligned}\frac{\partial X_1(j,t)}{\partial t} &= 1 - X_1(j,t) - Da_1 X_1(j,t) X_3^2(j,t) \\ &\quad + D_1[X_1(j+1,t) - 2X_1(j,t) + X_1(j-1,t)], \\ \frac{\partial X_2(j,t)}{\partial t} &= \beta - X_2(j,t) - Da_2 X_2(j,t) X_3^2(j,t) \\ &\quad + D_2[X_2(j+1,t) - 2X_2(j,t) + X_2(j-1,t)], \\ \frac{\partial X_3(j,t)}{\partial t} &= 1 - (1 + Da_3) X_3(j,t) + \alpha[Da_1 X_1(j,t) + Da_2 X_2(j,t)] X_3^2(j,t) \\ &\quad + D_3[X_3(j+1,t) - 2X_3(j,t) + X_3(j-1,t)], \qquad (3)\end{aligned}$$

with $D_i = d_i/(\Delta x)^2$, Δx being the spatial mesh size; $j = 1, 2, \ldots, N$. The number of degrees of freedom is significantly increased to $3N$ and the characterization of the system dynamics is, now, not a trivial task. In this study, the diffusion coefficients of the species A and D are assumed equal throughout and much greater than that for the species B, i.e., $D_1 = D_2 > D_3$. The unequal rates of diffusion and the local nonlinearity may then result in the formation of complex spatiotemporal patterns including spatiotemporal chaos depending on the parameter values. In figure 2b is shown the spatiotemporally chaotic dynamics exhibited by (3) for $N = 64$ and assuming periodic boundary conditions. Initially, all the N cells were assigned identical initial conditions and the parameter values were chosen corresponding to single cell (1) exhibiting chaotic dynamics (figure 1b). The extended system (3) was then simulated using the fourth order Runge–Kutta algorithm with time step $\Delta t = 0.0002$. For time $t < 40$ (in dimensionless units), the system dynamics is seen to be spatially correlated though temporally it is uncorrelated and chaotic. Note that the diffusion mechanism is not active in this region because of identical initial conditions assumed over the entire spatial domain, and therefore all the cells evolve synchronously. Now at time $t = 40$, a perturbation was given to the set of eleven cells lying in the central region of the lattice. This perturbation assumed that the dynamics of the chosen cells progressively went out of phase with each other by time $t = 0.002$. A spread of this perturbation to the system boundaries by diffusion, with a complete loss in the spatial correlation, is qualitatively seen in figure 2b and this dynamical behaviour was seen to persist indefinitely. Characterization of this loss in correlation and their effects, however, need to be studied quantitatively to make more meaningful interpretations. The following section discusses a means of doing so.

3. Characterization of reaction-diffusion dynamics

The characterization of spatiotemporal dynamics in terms of the Lyapunov exponents [52] which quantify the growth rate of trajectories for infinitesimally small perturbations will be discussed in this section. For the spatially discretized system (3) with three chemical species interacting on a one-dimensional lattice of size N, there are $3N$ dependent variables. Hence the complete Lyapunov spectrum would have $3N$ exponents and their calculation would be computationally very intensive, especially for large N. To alleviate this difficulty, in an earlier study we discussed the analysis of the Lyapunov spectrum for sub-systems of size $n_s (\ll N)$ and studied the behaviour of the system invariants as $n_s \to N$ [40]. For the sake of completeness in discussing issues related to the control of these systems, we shall now review this approach, highlighting the major results obtained.

For the extended system (3) with κ denoting the set of fixed parameters, the temporal evolution of the concentrations of each of the species for chosen values of D_i may be functionally written as

$$\frac{\mathrm{d}X_i(k,t)}{\mathrm{d}t} = F_{i,k}(X_1(k,t), \ldots, X_3(k,t), X_i(k \pm 1, t), \kappa), \tag{4}$$

where the first index $i = 1, 2, 3$ of the functionals $F_{i,k}$ denotes the respective species and the second index $k = 1, \ldots, n_s$ represents a sub-system lattice site. The Lyapunov exponents of the sub-system of size n_s may then be calculated in a manner similar to that for the full system [52], viz.,

$$\lambda_l^{(s)} = \lim_{t \to \infty} \sup \frac{1}{t} \ln \frac{|\, \delta \mathbf{X}^l(t) \,|}{|\, \delta \mathbf{X}^l(t_0) \,|}, \quad l = 1, \ldots, 3n_s, \tag{5}$$

where $\delta\mathbf{X}^l(t_0)$ refers to the infinitesimal perturbation from a reference state at $t = t_0$ and $\delta\mathbf{X}^l(t)$ its evolution at a time t. While calculating these exponents for (4), open boundary conditions were assumed at the sub-system boundary sites, i.e., at $k = 1$ and $k = n_\mathrm{s}$. The open boundary conditions indicate that only for the purposes of evaluating the λ_l's, the flow of information from the outer site of the sub-system boundaries, may be likened, to presence of noise in the analysis. Alternatively, for the 1d lattice, we may characterize the sub-system dynamics on excluding the outermost sub-system boundary sites, although time-series signals at these boundary sites are available. The effective sub-system size for characterization is now $n_\mathrm{s} - 2$, and the invariant measures of (4) can be studied accurately.

The characterization of the sub-system dynamics may now be carried out by applying the well-known Kaplan and Yorke (KY) conjecture [53] to the spectrum of sub-system Lyapunov exponents, $\lambda_l^{(\mathrm{s})}$. For doing so, we define the effective sub-system Lyapunov dimension, $d_L^{(\mathrm{s})}$, as

$$d_L^{(\mathrm{s})} = l + \frac{1}{|\lambda_{l+1}^{(\mathrm{s})}|}\sum_{i=1}^{l}\lambda_i^{(\mathrm{s})}, \tag{6}$$

where l is the largest integer for which the sum of the exponents, $\lambda_1^{(\mathrm{s})} + \cdots + \lambda_l^{(\mathrm{s})} \geq 0$. If $\lambda_1 < 0$, then $d_L^{(\mathrm{s})} = 0$ and if $l = 3n_\mathrm{s}$, then $d_L^{(\mathrm{s})} = 3n_\mathrm{s}$ [54]. Further reduction in the computational effort is possible as it may not be necessary to evaluate all the $3n_\mathrm{s}$ sub-system exponents. A relationship in the variation of $d_L^{(\mathrm{s})}$ with n_s would then help in establishing the number of effective variables required to determine the long-term behaviour of the complete system dynamics. From a knowledge of $d_L^{(\mathrm{s})}$ we can define the sub-system dimension density, $\rho^{(\mathrm{s})}$, as

$$\rho^{(\mathrm{s})} = \frac{d_L^{(\mathrm{s})}}{n_\mathrm{s}}, \tag{7}$$

by normalizing the sub-system Lyapunov dimension with its size n_s. The behaviour of $\rho^{(\mathrm{s})}$ as a function of n_s may then importantly help in identifying a critical sub-system size, n_sc, required for accurate characterization of the complete system dynamics.

Another important invariant measure is the Kolmogorov–Sinai (KS) entropy used for quantifying the mean information production and the growth of uncertainty in a system subjected to small perturbations [53]. The KS entropy is defined as the sum of the positive Lyapunov exponents λ_+ [52, 55]. For regular predictable behaviour, the KS entropy is zero while for chaotic systems it takes a finite positive value and tends to infinity for a stochastic process. In the present analysis we shall use this relationship for evaluating the sub-system KS-entropy, $h^{(\mathrm{s})}$, as

$$h^{(\mathrm{s})} = \sum \lambda_+^{(\mathrm{s})}, \tag{8}$$

and the normalized sub-system entropy, $\langle h^{(\mathrm{s})}\rangle$, as

$$\langle h^{(\mathrm{s})}\rangle = \sum \lambda_+^{(\mathrm{s})}/n_\mathrm{s}. \tag{9}$$

We now discuss the role of these sub-system invariants in the characterization of the full spatiotemporal system. In figure 3a is shown a plot of the sub-system dimension

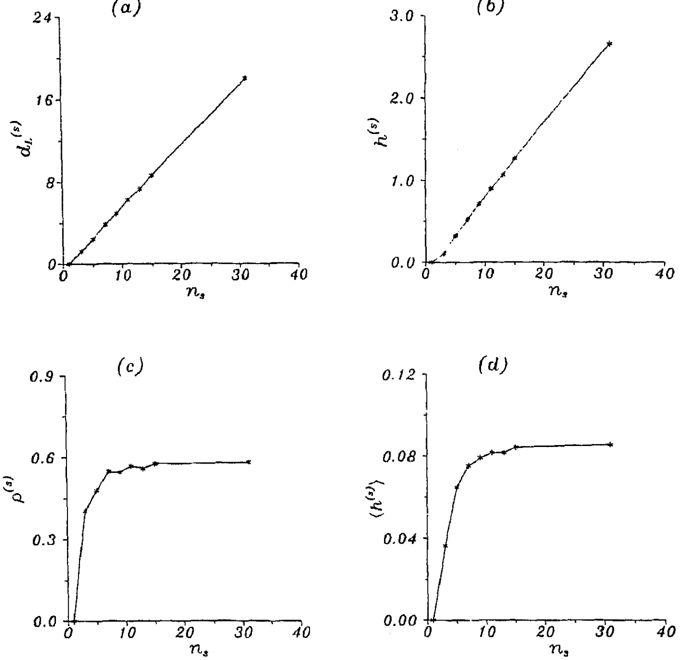


Figure 3. Behavior of the sub-system invariants as a function of its size n_s: (**a**) Lyapunov dimension, $d_L^{(s)}$; (**b**) entropy, $h^{(s)}$; (**c**) dimension density $\rho^{(s)}$; (**d**) normalized entropy, $\langle h^{(s)} \rangle$.

$d_L^{(s)}$ as a function of its size n_s. A clear linear scaling relationship in the sub-system dimension is seen for increasing sub-system size (figure 3a). This suggests that it may indeed be possible to determine the effective dimensionality of the whole system from an analysis of a relatively small sub-system. Some generic conjectures based on dimensionality have also been seen [34]. A clear linear scaling relationship with increasing sub-system size was also observed for the sub-system KS entropy, $h^{(s)}$, shown in figure 3b. The sub-system dimension density, $\rho^{(s)}$, is however, seen to saturate to a constant value for n_s greater than a critical sub-system size n_{sc} (figure 3c). This converging behaviour of $\rho^{(s)}$ for $n_s > n_{sc}$, may importantly help in assessing the number of degrees of freedom required for the characterization of the whole system. The saturating behaviour observed for the normalized sub-system entropy, $\langle h^{(s)} \rangle$, (figure 3d) suggests that even though the total entropy of the sub-system may increase with its size, the average rate of information loss/gain levels off for $n_s > n_{sc}$. These results suggest that sub-system analysis may suffice for the characterization of the full system dynamics. Similar scaling relationships in the system invariants as a function of sub-system size have also been observed for coupled maps with near neighbour interactions indicating possible generalizations [35, 36]. The knowledge of these sub-system properties may be used in identifying favourable conditions for controlling the spatiotemporal system and in the following section we shall further elaborate on this aspect.

4. Synchronization and control of spatiotemporal chaos

In this section, we discuss the synchronization and control of the spatiotemporally chaotic dynamics of a response system driven by monitored time-series signals from a spatiotemporal process for various objectives like stabilizing the chaotic behaviour of the response system and its ability to mimic the process dynamics under different situations [16]. In practice, there exists constraints in the monitoring of scalar time-series data from the entire spatial domain of the process, and moreover, this data may generally be noisy due to errors arising during measurement. It would be desirable to consider the control of the spatiotemporal dynamics of replica and non-replica response systems for these limitations.

4.1 *Synchronization in replica systems*

To begin with, we first consider the response system to be a replica of a spatiotemporally chaotic process, i.e., the response and the process systems are assumed to be identical in all respects except in their initial conditions, $X_i(j,0) \neq \hat{X}_i(j,0)$, where $\hat{X}_i(j,0)$ denote the response system variables. The response system model may then be written in its variables, $\hat{X}_i(j,t)$, as

$$\begin{aligned}
\dot{\hat{X}}_1(j,t) &= F_1(\hat{X}_1(j,t),\hat{X}_3(j,t)) \\
&= 1-\hat{X}_1(j,t) - Da_1\hat{X}_1(j,t)\hat{X}_3^2(j,t) \\
&\quad + D_1[\hat{X}_1(j+1,t) - 2\hat{X}_1(j,t) + \hat{X}_1(j-1,t)], \\
\dot{\hat{X}}_2(j,t) &= F_2(\hat{X}_2(j,t),\hat{X}_3(j,t)) \\
&= \beta-\hat{X}_2(j,t) - Da_2\hat{X}_2(j,t)\hat{X}_3^2(j,t) \\
&\quad + D_2[\hat{X}_2(j+1,t) - 2\hat{X}_2(j,t) + \hat{X}_2(j-1,t)], \\
\dot{\hat{X}}_3(j,t) &= F_3(\hat{X}_1(j,t),\hat{X}_2(j,t),\hat{X}_3(j,t)) \\
&= 1-(1+Da_3)\hat{X}_3(j,t) + \alpha[Da_1\hat{X}_1(j,t) + Da_2\hat{X}_2(j,t)]\hat{X}_3^2(j,t) \\
&\quad + D_3[\hat{X}_3(j+1,t) - 2\hat{X}_3(j,t) + \hat{X}_3(j-1,t)],
\end{aligned} \tag{10}$$

where, $j = 1,2,\ldots,N$. If the response system (10) is now assumed to be driven by scalar time-series signals, $X_3(j,t)$, $j = 1,2,\ldots,N$, from the process, its dynamics follow

$$\begin{aligned}
\dot{\hat{X}}_1(k,t) &= F_1(\hat{X}_1(k,t),X_3(k,t)), \\
\dot{\hat{X}}_2(k,t) &= F_2(\hat{X}_2(k,t),X_3(k,t)), \\
\dot{\hat{X}}_3(k,t) &= F_3(\hat{X}_1(k,t),\hat{X}_2(k,t),X_3(j,t)),
\end{aligned} \tag{11}$$

and was found to completely synchronize with the process dynamics, in all the $3N$ variables. This is clearly seen in figure 4a from the diminishing behaviour of the relative error in the response and process dynamics, i.e., $e_2'(j,t) = [\hat{X}_2(j,t) - X_2(j,t)]/X_2(j,t) \to 0$. (Similar synchronized behaviour was observed in the other two variables also).

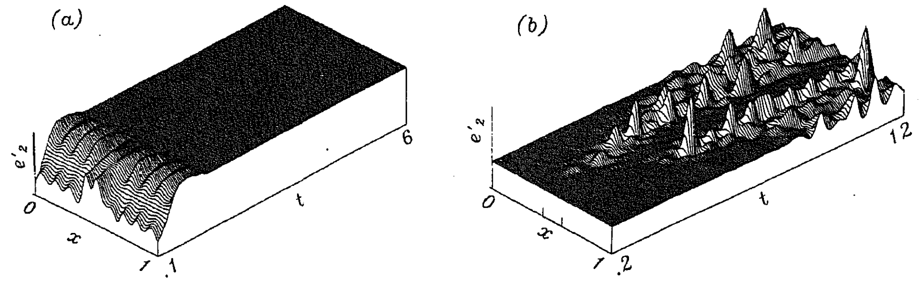


Figure 4. Synchronization in a replica system for parameter values in figure 2. The relative space-time error behaviour for $e'_2(j,t)$ shown. (**a**) Complete synchronization in the spatio-temporally chaotic dynamics of the response system obtained on using driving signals, $X_3(j,t)$, $j=1,\dots,N$, from the process; e'_2 axis-scale: $(-0.005, 0.0007)$. (**b**) On using sub-system driving signals, $X_3(k,t)$, $k=1,\dots,n_s$, for $n_s=11$, from the central region (shown by tics), synchronization was observed only within the sub-system. e'_2 axis-scale: $(-1.11, 6.35)$.

We now show that it may be possible to explain the observed synchronization behaviour of the combined response-process dynamics using Lyapunov stability analysis. The method suggests that a given system is stable, if any continuously derivable positive definite function, L, (called the Lyapunov function) can be defined along a trajectory with the property that its time derivative, $\dot{L} \leq 0$, as $t \to \infty$ [56]. This method may be advantageously used since it affords analytical reasoning of the stability properties of the system. In the present study, we show by constructing suitable Lyapunov functionals in terms of error variables, $e_i(j,t) = \hat{X}_i(j,t) - X_i(j,t)$, that synchronization in the dynamics between the response and the process may be assessed. The dynamical equations for the error function variables $e_i(j,t)$ are

$$\begin{aligned}\dot{e}_1(j,t) &= \dot{\hat{X}}_1(j,t) - \dot{X}_1(j,t)\\ &= -e_1(j,t)[1 + Da_1X_3^2(j,t)]\\ &\quad + D_1[e_1(j-1,t) - 2e_1(j,t) + e_1(j+1,t)], \qquad (12)\end{aligned}$$

$$\begin{aligned}\dot{e}_2(j,t) &= \dot{\hat{X}}_2(j,t) - \dot{X}_2(j,t)\\ &= -e_2(j,t)[1 + Da_2X_3^2(j,t)]\\ &\quad + D_2[e_2(j-1,t) - 2e_2(j,t) + e_2(j+1,t)], \qquad (13)\end{aligned}$$

$$\begin{aligned}\dot{e}_3(j,t) &= \dot{\hat{X}}_3(j,t) - \dot{X}_3(j,t)\\ &= \alpha[Da_1e_1(j,t) + Da_2e_2(j,t)]X_3^2(j,t), \qquad (14)\end{aligned}$$

written using (3) and (10). Then, for a chosen positive Lyapunov functional L defined as $L = \sum_i \sum_j e_i^2(j,t)$, its derivative $\dot{L}$ is

$$(1/2)\dot{L} = \sum_{i=1}^{3}\sum_{j=1}^{N} e_i(j,t)\dot{e}_i(j,t). \qquad (15)$$

Substituting (12–14) in (15) and for the assumed periodic boundary conditions, we obtain on simplification

$$\sum_{j=1}^{N} e_1(j,t)\dot{e}_1(j,t) = -\sum_{j=1}^{N} e_1{}^2(j,t)[1 + Da_1 X_3^2(j,t)]$$
$$- D_1 \sum_{j=1}^{N} [e_1(j,t) - e_1(j+1,t)]^2, \tag{16}$$

$$\sum_{j=1}^{N} e_2(j,t)\dot{e}_2(j,t) = -\sum_{j=1}^{N} e_2{}^2(j,t)[1 + Da_2 X_3^2(j,t)]$$
$$- D_2 \sum_{j=1}^{N} [e_2(j,t) - e_2(j+1,t)]^2, \tag{17}$$

$$\sum_{j=1}^{N} e_3(j,t)\dot{e}_3(j,t) = \alpha \sum_{j=1}^{N} \{e_3(j,t)[Da_1 e_1(j,t) + Da_2 e_2(j,t)]X_3^2(j,t)\}. \tag{18}$$

It may be noted from (12) and (13) that the error dynamics of $e_1(j,t)$ and $e_2(j,t)$ are dependent only on their respective error function variables, and on the square of the driving signals, $X_3(j,t)$. Equations (12) and (13) are decoupled in this sense and may be considered independently for assessing the synchronization behaviour of $\hat{X}_1$ and $\hat{X}_2$ by independent Lyapunov functions $L_i = \sum_j e_i^2(j,t)$, $i = 1,2$. Then, from (16) and (17), it is clear that the $\dot{L}_i$'s are negative definite, and $e_1(j,t), e_2(j,t) \to 0$ as $t \to \infty$ by the Lyapunov stability theorem. Using these conditions in (15), $\dot{L}$ is strictly nonpositive and indicates the stable convergence of the error dynamics (12–14) to the origin. That is, the response system would dynamically synchronize with the process dynamics in all its variables.

On the other hand, if the time-series signals are available from a sub-system of size $n_s (\ll N)$, then following the above Lyapunov stability analysis, synchronization should be possible only within the sub-system region. This is because for the variables outside the sub-system, the error function variables are governed by

$$\dot{e}_1(l,t) = -e_1(l,t) - Da_1[\hat{X}_1(l,t)\hat{X}_3^2(l,t) - X_1(l,t)X_3^2(l,t)]$$
$$+ D_1[e_1(l-1,t) - 2e_1(l,t) + e_1(l+1,t)], \tag{19}$$
$$\dot{e}_2(l,t) = -e_2(l,t) - Da_2[\hat{X}_2(l,t)\hat{X}_3^2(l,t) - X_2(l,t)X_3^2(l,t)]$$
$$+ D_2[e_2(l-1,t) - 2e_2(l,t) + e_2(l+1,t)], \tag{20}$$
$$\begin{aligned}\dot{e}_3(l,t) = {} & -(1 + Da_3)[\hat{X}_3(l,t) - X_3(l,t)] \\ & + \alpha Da_1[\hat{X}_1(l,t)\hat{X}_3^2(l,t) - X_1(l,t)X_3^2(l,t)] \\ & + \alpha Da_2[\hat{X}_2(l,t)\hat{X}_3^2(l,t) - X_2(l,t)X_3^2(l,t)] \\ & + D_3[e_3(l-1,t) - 2e_3(l,t) + e_3(l+1,t)],\end{aligned} \tag{21}$$

where l denotes the lattice sites outside the sub-system, i.e., $l = 1, 2, \ldots, N$, excluding the n_s sub-system lattices sites from the central region. Note that the terms (19) and (20) are now not necessarily negative and synchronization in these variables cannot be guaranteed.

This behaviour was numerically observed on driving the response system (10) by scalar time-series signals from a sub-system of the process of size $n_s = 11$, from the central region of the lattice. The space-time plot of the error function $e'_2(j,t)$ in figure 4b clearly shows only sub-system synchronization.

In situations when a suitable Lyapunov functional cannot be constructed, the synchronization capability of the response system with that of the process may be inferred from an analysis of the conditional Lyapunov exponents of the response system [18, 19]. Following the results of the sub-system invariants discussed in § 3, it may be sufficient to analyse the conditional Lyapunov exponents of only a sub-system of size $n_s(> n_{sc})$ of the response system. The procedure for calculating the conditional Lyapunov exponents of a sub-system is similar to that of the sub-system Lyapunov exponents discussed in the § 3, but now we need to monitor the growth rate of only $2n_s$ sets of orthonormal vectors ($\delta X_i(k,t)$, $i = 1,2$; $k = 1,2,\ldots,n_s$) in a linearized region of (11). The maximum conditional Lyapunov exponent for a sub-system of size $n_s = 11$ of the response system (11) was found to be negative ($\lambda_{\max} \sim -0.07$), as expected. Thus, the sub-system conditional exponents may be used to assess the synchronization capability of the response system (see figure 4a).

4.2 *Synchronization in non-replica systems*

Next, we consider the situation when the response system is a non-replica of the process, i.e., the response and the process are assumed to be operating in different parametric regimes. Synchronization of the response dynamics with that of the process in this situation may be difficult because of inaccuracies in its parameter values. For example, consider both the process and the response system exhibiting spatiotemporal chaos, but with a chosen control parameter, β, of the process having a value 2.93 (which may not be known *a priori* in a given experimental situation), and the corresponding response parameter, $\hat{\beta}$, set incorrectly at 2.73 at $t = 0$. The hence response and the process evolve asynchronously because of this parametric inaccuracy, and simple driving as employed in subsection 4.1 for replica systems, was found to be insufficient for the complete dynamical synchronization. There exists a need, therefore, to correctly estimate the parametric inaccuracy. To do so, we focus attention on a convenient strategy which allows to correctly estimate the inaccuracy in the control parameter via suitably chosen functional forms for parametric self-adaptation. For low-dimensional chaotic systems it has been shown that many simple choices of the functional forms may indeed be used for parametric self-adaptation and control [57–65].

For parametric self-adaptation of spatiotemporal systems, we begin by introducing space-time dependence in the chosen response parameter, i.e., $\hat{\beta}(j,t)$ and define a new variable, $\Delta\beta(j,t)$, denoting the time-dependent parametric corrections to be made such that, $\hat{\beta}(j,t) = [\hat{\beta}(j,0) + \Delta\beta(j,t)] \to \beta$ as $t \to \infty$. For convenience, it was assumed that at $t = 0$, $\hat{\beta}(j,0) = \hat{\beta}$, and $\Delta\beta(j,0) = 0$, for $j = 1,2,\ldots,N$ with the following simple functional form for the dynamical corrections, $\Delta\beta(j,t)$, viz.,

$$\frac{\mathrm{d}\Delta\beta(j,t)}{\mathrm{d}t} = -[\hat{X}_3(j,t) - X_3(j,t)] = -e_3(j,t). \tag{22}$$

It would be interesting to see if a Lyapunov functional can predict the stability of the error dynamics with the parametric self-adapter equation (22). The error dynamics in this case follows

$$\dot{e}_1(j,t) = -e_1(j,t)[1 + Da_1 X_3^2(j,t)] + D_1[e_1(j-1,t) - 2e_1(j,t) + e_1(j+1,t)], \tag{23}$$

$$\dot{e}_2(j,t) = (\hat{\beta}(j,0) + \Delta\beta(j,t)) - \beta - e_2(j,t)[1 + Da_2 X_3^2(j,t)] + D_2[e_2(j-1,t) - 2e_2(j,t) + e_2(j+1,t)], \tag{24}$$

$$\dot{e}_3(j,t) = \alpha[Da_1 e_1(j,t) + Da_2 e_2(j,t)]X_3^2(j,t). \tag{25}$$

It may be noted that, similar to the analysis for the replica system, the error dynamics of $e_1(j,t)$ is again dependent only on $e_1(j,t)$ and the square of the time-series signals, $X_3^2(j,t)$. That is, we can define a Lyapunov function $L_1 = \sum_j e_1^2(j,t)$ such that

$$(1/2)\dot{L}_1 = \sum_{j=1}^{N} e_1(j,t)\dot{e}_1(j,t) = -\sum_{j=1}^{N} e_1^2(j,t)[1 + Da_1 X_3^2(j,t)] - D_1 \sum_{j=1}^{N} [e_1(j,t) - e_1(j+1,t)]^2 \tag{26}$$

is negative definite. Hence, following the Lyapunov stability theorem, $e_1(j,t) \to 0$, irrespective of the dynamical behaviour of the time-series signals, $X_3(j,t)$ and the remaining variables of the response system. This implies that synchronization in $\hat{X}_1(j,t)$ is guaranteed and this feature may be used for studying the synchronization behaviour with respect to the remaining variables. Thus, on defining a positive Lyapunov functional of the form $L = \sum_j e_1^2(j,t) + e_1^2(j,t)[e_2^2(j,t) + e_3^2(j,t) + \Delta\beta^2(j,t)]$, we find that its derivative, $\dot{L}$, is strictly nonpositive as $t \to \infty$. Hence, dynamical synchronization of the response system with that of the process is likely with the corrections, $\Delta\beta(j,t)$, appropriately estimated. A numerical study confirmed the above prediction (results not shown).

It is interesting to note that the above Lyapunov functional analysis is valid even if the monitored time-series signals are noisy due to measurement errors. That is, we assume the response model to be driven by noisy scalar time-series data, $X_3'(j,t)$, of the form

$$X_3'(j,t) = X_3(j,t) + \gamma\eta(j,t), \quad j = 1,\ldots,N, \tag{27}$$

where γ is the intensity of Gaussian noise, $\eta(j,t)$. The inherent tendency of the response to reduce the errors $e_i(j,t)$, $i = 1,2,3$ still persists with a noise reduction capability depending on the noise level and frequency. Our studies showed that when the driving signals $X_3'(j,t)$, were assumed to be available from all the lattice sites, synchronization in the response variables was observed over the entire spatial domain even for noisy time-series signals. This is depicted in figures 5a and b, by the diminishing behaviour of the relative error signals $e_i'(j,t) = [\hat{X}_i(j,t) - X_i(j,t)]/X_i(j,t)$, $i = 1,2$, with the convergence of $\Delta\beta(j,t)$ to the mean correction, $\beta - \hat{\beta}(j,0) = 0.2$, shown in figure 5c. Note that $\Delta\beta(j,t)$ fluctuates around this mean correction in an attempt to filter the effects of noise on the dynamics of the response system. In keeping with the observations made earlier on studying synchronization properties for data monitored from a subsystem, the results obtained with noisy driving signals from a sub-system, i.e., $X_3'(k,t)$,

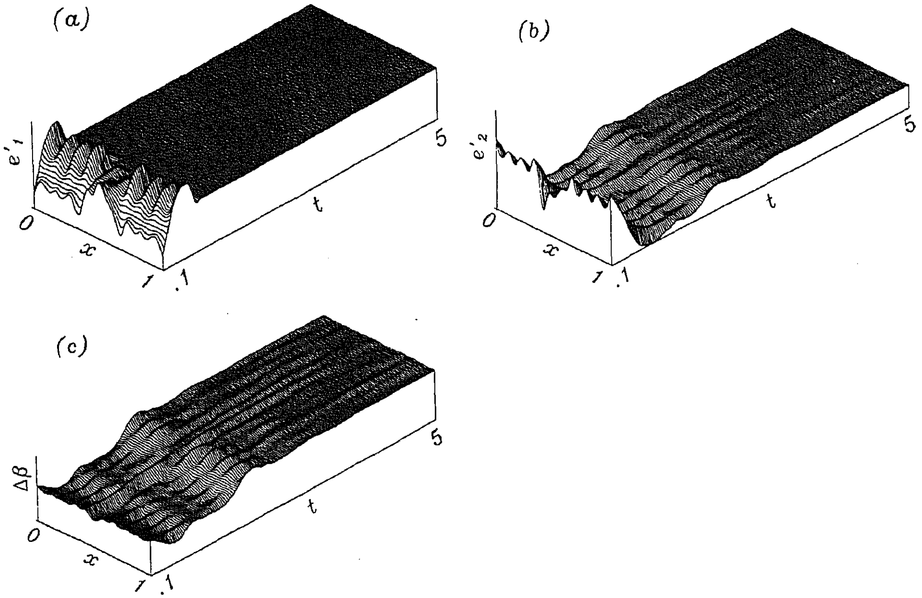


Figure 5. Synchronization in a non-replica system for an initial inaccuracy in response parameter $\hat{\beta}$, i.e., $\hat{\beta} = 2.73$, while process parameter $\beta = 2.93$. The study assumed the response system to be driven by time-series signals, $X_3'(j,t)$, $j = 1, 2, \ldots, N$, corrupted with measurement noise of intensity $\gamma = 0.00002$. **(a, b)** The relative error signals, $e_1(j,t), e_2(j,t) \rightarrow 0$, showing the synchronization of the response dynamics with the process; e_1' axis-scale: $(-0.051, 0.316)$; e_2' axis-scale: $(-0.051, 0.316)$. **(c)** Space-time behaviour of the correction, $\Delta\beta(j,t)$, converging accurately to 0.2 over the entire spatial domain is seen; $\Delta\beta$ axis-scale: $(-0.051, 0.251)$.

$k = 1, 2, \ldots, n_s$ showed that synchronization was again possible only within the sub-system region. This is shown in figures 6a and b, where the relative synchronization error, $e_i'(j,t)$, $i = 1, 2$, has been plotted. Note that for the lattice sites outside the sub-system (and not covered by the adapter equations (22)), an average adaptation $\langle\Delta\beta\rangle = \sum_k \Delta\beta(k,t)/n_s$ was employed. The dynamical convergence of this average correction, $\langle\Delta\beta\rangle$, to 0.2 is shown in figure 6c. These results suggest that even with limited and inaccurate sub-system information, it may be possible to correctly estimate the inaccuracy in the chosen control parameter, though, complete dynamical synchronization of the response in all its variables may be a difficult task.

4.3 *Stabilization of spatiotemporal chaos*

Frequently, it may be necessary to stabilize the spatiotemporally chaotic dynamics and in this subsection we present the results obtained on driving a response system using limited and noisy time-series signals from a process exhibiting time-independent stable dynamics. The response system was assumed to exhibit spatiotemporal chaos with $\hat{D}a_3 = 80$ and the aim was to self-adapt this parameter so that the response system

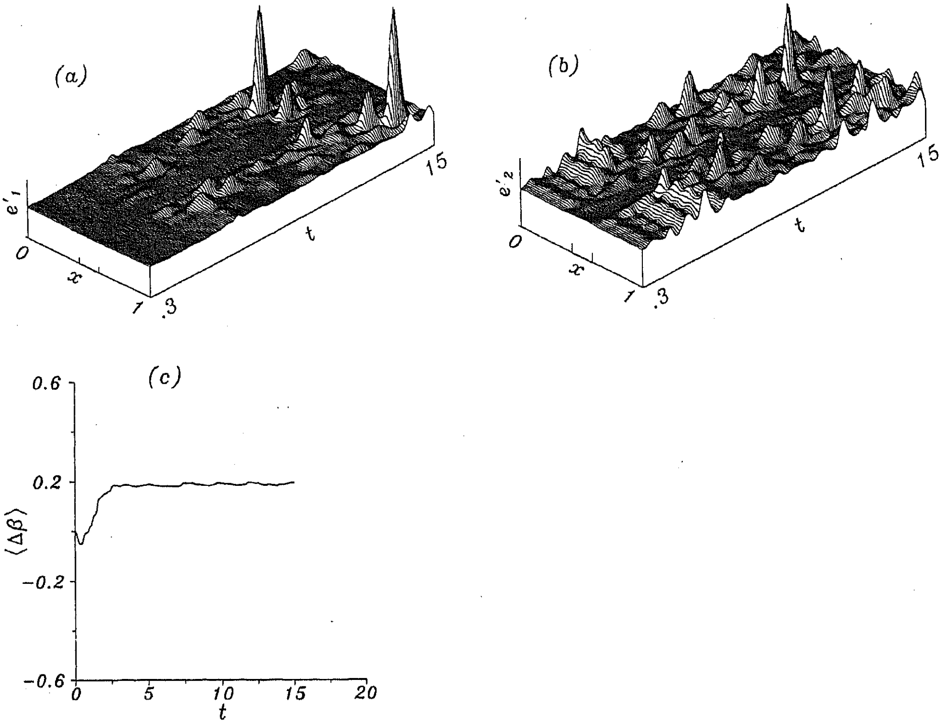


Figure 6. Synchronization behaviour of a spatiotemporally chaotic dynamics of the response when driven with noisy data from a sub-system of size $n_s = 11$, from the central region of the lattice. **(a, b)** Synchronization observed only within the sub-system region where, $e_1(j,t), e_2(j,t) \to 0$; e'_1 axis-scale: $(-24.89, 182.8)$; e'_2 axis-scale: $(-1.117, 8.316)$. **(c)** The temporal behaviour of $\langle \Delta\beta \rangle$, is seen to accurately estimate the correction required.

a priori). Thus, in this case, we study the inverse problem of locating the control parameter value which yields a desired output in the dependent variables of the response system. The dynamical equations for the error functions $e_i(j,t)$, in this situation follows

$$\begin{aligned} \dot{e}_1(j,t) &= -e_1(j,t)[1 + Da_1 X_3^2(j,t)] \\ &\quad + D_1[e_1(j-1,t) - 2e_1(j,t) + e_1(j+1,t)], \end{aligned} \tag{28}$$

$$\begin{aligned} \dot{e}_2(j,t) &= -e_2(j,t)[1 + Da_2 X_3^2(j,t)] \\ &\quad + D_2[e_2(j-1,t) - 2e_2(j,t) + e_2(j+1,t)], \end{aligned} \tag{29}$$

$$\begin{aligned} \dot{e}_3(j,t) &= -\Delta Da_3(j,t) X_3(j,t) \\ &\quad + \alpha[Da_1 e_1(j,t) + Da_2 e_2(j,t)] X_3^2(j,t), \end{aligned} \tag{30}$$

on introducing space-time dependence in the chosen response parameter, i.e., $\hat{D}a_3(j,t)$ with $Da_3(j,0) = \hat{D}a_3$. As before, we now define a new variable, $\Delta Da_3(j,t)$, for the parametric corrections required to be made to the response parameter $\hat{D}a_3$ such that the

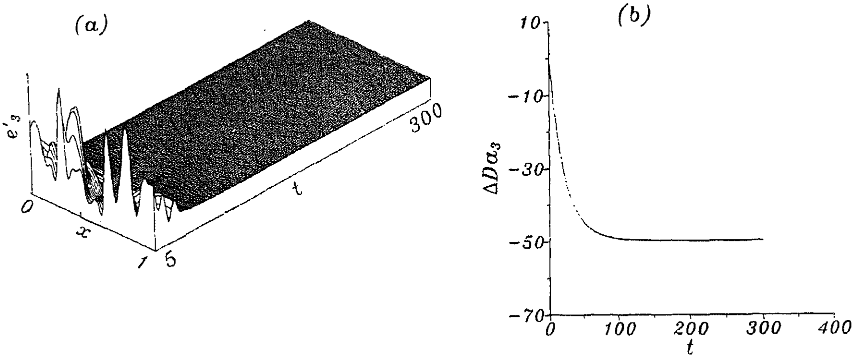


Figure 7. Stabilization of the spatiotemporal chaos with noise reduction. (**a**) The relative error $e'_3(j,t) \to 0$ indicating stabilization of the response system at the desired state; e'_3 axis-scale: (−2.88, 27.89). (**b**) The convergence of $\langle \Delta Da_3 \rangle$ to the initial difference $Da_3 - \hat{D}a_3 = -50$.

true process value Da_3 is realized as $t \to \infty$. For convenience, we assign at $t = 0$, $\Delta Da_3(j,0) = 0$ and to bring out the flexibility in choosing a dynamical form for the parametric self-adaptation, we now consider

$$\frac{\mathrm{d}\Delta Da_3(j,t)}{\mathrm{d}t} = [\hat{X}_3(j,t) - X_3(j,t)]X_3(j,t) = e_3(j,t)X_3(j,t). \tag{31}$$

Then, for a positive Lyapunov functional L defined as

$$L = \sum_i \sum_j e_i^2(j,t) + \sum_j \Delta Da_3^2(j,t), \tag{32}$$

its derivative, $\dot{L}$, is again seen to be strictly nonpositive and from the Lyapunov stability theorem it follows that the response system would completely synchronize with the stable process dynamics for appropriate dynamic corrections, $\Delta Da_3(j,t)$. This was also numerically confirmed for noisy driving signals, $X'_3(k,t)$, from a sub-system of the process. However, unlike in the previous studies, we now observe synchronization over the entire spatial domain. This is clearly seen in figure 7a for the relative error signal, $e'_3(j,t) \to 0, j = 1, \ldots, N$, when the response system was assumed to be driven by time-series signals from a sub-system of size $n_s = 11$, from the central region of the process (following (27)), viz., $X'_3(k,t)$, $k = 1, 2, \ldots, n_s$. As in the previous study, for the lattice sites outside the sub-system (and not covered by (31)), an average adaptation $\langle \Delta Da_3 \rangle = \sum_k \Delta Da_3(k,t)/n_s$ was employed. The convergence of $\langle \Delta Da_3 \rangle$ to the initial difference in the response and process parameters, $Da_3 - \hat{D}a_3 = -50$, is seen in figure 7b. Similar results were observed on varying the noise strength γ. Thus, the self-adaptation mechanism can be effectively used even in the presence of reasonable extents of noise. These results also suggest that it may be possible to stabilize the spatiotemporal chaos in real-life experiments, using only scalar driving signals from a subsystem of an experimental process along with an appropriate self-adaptation mechanism as controller feedback. Importantly, a knowledge of the nonlinear mathematical model or a process

In the different situations discussed above, inaccuracies were assumed in only a single system parameter. However in many situations, the inaccuracies may exist in more than one parameter and for synchronization to occur, simultaneous correction in all the inaccurate parameters would then be required. This topic is being actively pursued and different strategies, e.g., parametric self-adaptation [64, 66], optimization [68], neural networks [69, 70], etc. have been reported. In this subsection, we show that multi-parameter control of a spatiotemporally chaotic system via judicious self-adaptation is feasible and advantageous because of its simplicity. For the sake of convenience, we rewrite the process equations as

$$\begin{aligned}\dot{X}_1(j,t) =& 1 - X_1(j,t) - Da_1X_1(j,t)X_3^2(j,t)\\ &+ D_1[X_1(j+1,t) - 2X_1(j,t) + X_1(j-1,t)],\\ \dot{X}_2(j,t) =& \beta - X_2(j,t) - Da_2X_2(j,t)X_3^2(j,t)\\ &+ D_2[X_2(j+1,t) - 2X_2(j,t) + X_2(j-1,t)],\\ \dot{X}_3(j,t) =& A_1 - A_2X_3(j,t) + A_3[Da_1X_1(j,t) + Da_2X_2(j,t)]X_3^2(j,t)\\ &+ D_3[X_3(j+1,t) - 2X_3(j,t) + X_3(j-1,t)],\end{aligned} \tag{33}$$

where A_1, A_2, and A_3 are the chosen control parameters and may be identified as $A_1 = 1.0$, $A_2 = 1 + Da_3 = 81$, and $A_3 = \alpha = 1.5$ in (3). Let us assume that these are incorrectly set to $\hat{A}_1 = 1.1$, $\hat{A}_2 = 82$, and $\hat{A}_3 = 1.7$, respectively, in the response system driven by the time-series signals, $X_3(j,t), j = 1, 2, \ldots, N$. Thus,

$$\begin{aligned}\dot{\hat{X}}_1(j,t) =& 1 - \hat{X}_1(j,t) - Da_1\hat{X}_1(j,t)X_3^2(j,t)\\ &+ D_1[\hat{X}_1(j+1,t) - 2\hat{X}_1(j,t) + \hat{X}_1(j-1,t)],\\ \dot{\hat{X}}_2(j,t) =& \beta - \hat{X}_2(j,t) - Da_2\hat{X}_2(j,t)X_3^2(j,t)\\ &+ D_2[\hat{X}_2(j+1,t) - 2\hat{X}_2(j,t) + \hat{X}_2(j-1,t)],\\ \dot{\hat{X}}_3(j,t) =& \hat{A}_1 - \hat{A}_2X_3(j,t) + \hat{A}_3[Da_1\hat{X}_1(j,t) + Da_2\hat{X}_2(j,t)]X_3^2(j,t)\\ &+ D_3[X_3(j+1,t) - 2X_3(j,t) + X_3(j-1,t)].\end{aligned} \tag{34}$$

The aim is then to control the response system dynamics by proper estimation of the parametric corrections, viz., ΔA_1, ΔA_2 and ΔA_3. We first show that for an appropriate choice of driving signals and functional forms for self-adaptation, a Lyapunov functional may be formulated to assess the synchronization and control capability of the response system. The dynamical behaviour of the error functions, $e_i(j,t) = \hat{X}_i(j,t) - X_i(j,t)$, is

$$\begin{aligned}\dot{e}_1(j,t) =& -e_1(j,t)[1 + Da_1X_3^2(j,t)]\\ &+ D_1[e_1(j-1,t) - 2e_1(j,t) + e_1(j+1,t)],\end{aligned} \tag{35}$$

$$\begin{aligned}\dot{e}_2(j,t) =& -e_2(j,t)[1 + Da_2X_3^2(j,t)]\\ &+ D_2[e_2(j-1,t) - 2e_2(j,t) + e_2(j+1,t)],\end{aligned} \tag{36}$$

$$\begin{aligned}\dot{e}_3(j,t) = {}& \Delta A_1(j,t) - \Delta A_2(j,t)X_3(j,t)\\ &+ Da_1[\Delta A_3(j,t)\hat{X}_1(j,t) + A_3 e_1(j,t)]X_3^2(j,t)\\ &+ Da_2[\Delta A_3(j,t)\hat{X}_2(j,t) + A_3 e_2(j,t)]X_3^2(j,t),\end{aligned} \tag{37}$$

with the dynamical equations for the parametric corrections, $\Delta A_i(j,t)$, $i = 1, 2, 3$, chosen respectively, as

$$\Delta\dot{A}_1(j,t) = -[\hat{X}_3(j,t) - X_3(j,t)] = -e_3(j,t), \tag{38}$$

$$\Delta\dot{A}_2(j,t) = [\hat{X}_3(j,t) - X_3(j,t)]X_3(j,t) = e_3(j,t)X_3(j,t), \tag{39}$$

$$\begin{aligned}\Delta\dot{A}_3(j,t) &= -Da_1\hat{X}_1(j,t)[\hat{X}_3(j,t) - X_3(j,t)]X_3^2(j,t)\\ &\quad - Da_2\hat{X}_2(j,t)[\hat{X}_3(j,t) - X_3(j,t)]X_3^2(j,t)\\ &= [Da_1\hat{X}_1(j,t) + Da_2\hat{X}_2(j,t)]e_3(j,t)X_3^2(j,t).\end{aligned} \tag{40}$$

Then the derivative of a positive Lyapunov functional, $L = \sum_i \sum_j [e_i^2(j,t) + \Delta A_i^2(j,t)]$, is

$$(1/2)\dot{L} = \sum_i \sum_j [e_i(j,t)\dot{e}_i(j,t) + \Delta A_i(j,t)\Delta\dot{A}_i(j,t)], \tag{41}$$

that is,

$$\begin{aligned}(1/2)\dot{L} = &-\sum_{j=1}^{N}\{e_1^2(j,t)[1 + Da_1 X_3^2(j,t)] - D_1[e_1(j,t) - e_1(j+1,t)]^2\}\\ &-\sum_{j=1}^{N}\{e_2^2(j,t)[1 + Da_2 X_3^2(j,t)] - D_2[e_2(j,t) - e_2(j+1,t)]^2\}\\ &+\sum_{j=1}^{N} A_3 e_3(j,t)[Da_1 e_1(j,t) + Da_2 e_2(j,t)]X_3^2(j,t).\end{aligned} \tag{42}$$

It may be noted that (35) and (36) are again decoupled systems, and following the arguments presented in the previous subsections, we obtain $e_i(j,t) \to 0$ as $t \to \infty$, for $i = 1, 2$. Using these results in (41), it is clear that $\dot{L}$ is non-positive and dynamical synchronization of the response with the process should be possible. This was confirmed numerically and below we present our results obtained on using driving signals only from a sub-system of the process, i.e., in $X_3(k,t)$, $k = 1, 2, \ldots, n_s$. As before, we begin by introducing space-time dependence in the chosen control parameters, and assume that at $t = 0$, $\hat{A}_i(j,0) = \hat{A}_i \neq A_i$. The parametric corrections, $\Delta A_i(k,t)$'s, within the sub-system are assumed to be governed by (38–40). For the lattice sites outside the sub-system, an average adaptation, $\langle \Delta A_i \rangle = \sum_k \Delta A_i(k,t)/n_s$, were implemented. Figures 8a, 8b and 8c respectively show the convergence of $\langle \Delta A_1 \rangle \to -0.1$, $\langle \Delta A_2 \rangle \to -1.0$ and $\langle \Delta A_3 \rangle \to -0.2$, the required corrections. Thus, these results suggest that it may be possible to simultaneously adapt more than one system parameter using limited scalar time-series data from a sub-system of the process for appropriate choices of the functional forms for parametric corrections. It may be clarified that when $n_s = N$, not only adaptation of the parameters but control via synchronization of the complete space-time dynamics of the response system may be possible.

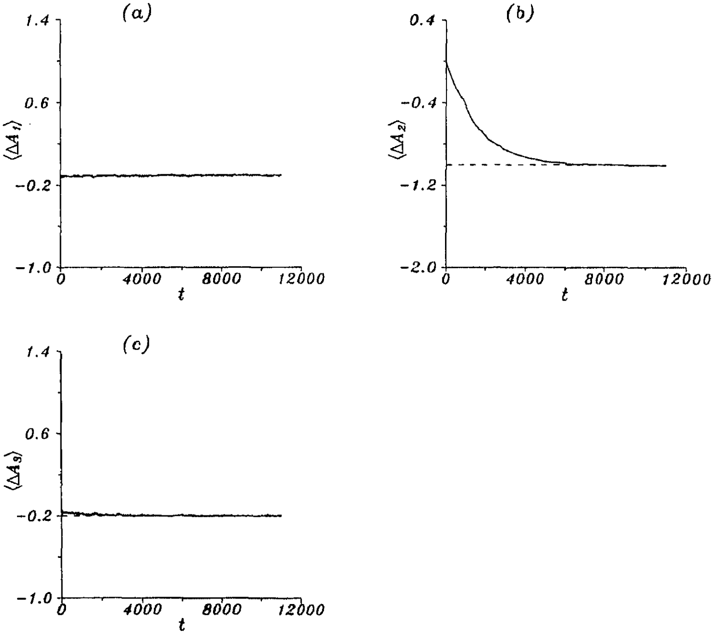


Figure 8. (**a, b, c**) Multi-parameter control via self-adaptation showing the accurate estimation of the corrections, $\langle \Delta A_i \rangle$, $i = 1, 2, 3$.

5. Conclusion

The analysis of sub-system dynamics for a nonlinear autocatalytic reaction-diffusion system as a function of increasing sub-system size showed interesting linear scaling relationships with respect to the invariant measures like the Lyapunov dimension and KS-entropy, while the dimension densities and normalized entropy were found to become independent of the sub-system size beyond a certain critical size n_{sc}. These results suggest that characterization of the complete system dynamics may be possible from an analysis of the Lyapunov spectrum of its sub-system of size $n_{\mathrm{s}} > n_{\mathrm{sc}}$ with considerable reduction in the computational effort. Our preliminary results for the autocatalytic reaction-diffusion system on a two-dimensional lattice has also shown similar trends for the system invariants as a function of its sub-system size. Thus, the approach of analyzing sub-system behaviour may be advantageously used in the characterization of more complex higher-dimensional systems. It may be noted that the studies discussed in this paper were carried out for nearest-neighbor diffusive coupling. However, in most practical situations, other transport mechanisms such as convection may also concurrently occur. It would then be worthwhile to study the validity of the scaling relationships in the sub-system invariant properties for systems involving more complex and long-range coupling.

We have used the above characterization results to analyse the synchronization and controllability of spatiotemporal systems exhibiting chaotic dynamics. The control objectives were carried out for inaccurate knowledge of the initial conditions, parameter settings, and limitations in the monitoring of time-series data from the spatial domain which may additionally be corrupted with noise. Our results using sub-system information show that it may be possible to correctly estimate the inaccuracies in the control parameter(s) via suitably chosen forms of parametric self-adaptation(s). Suppression of chaos and stabilization of the spatiotemporally chaotic dynamics to a desired stable state was also possible using sub-system driving signals even in the presence of reasonable extents of noise. However, for the synchronization of complex and chaotic dynamics, time-series data from the entire spatial domain may be required. Further, it has been shown that it may be possible to assess the synchronization capability of a spatiotemporal system from the stability analysis of suitably constructed Lyapunov functionals. This approach offers the advantage of *a priori* knowledge of the synchronization behaviour under different situations.

Acknowledgements

The authors gratefully acknowledge the financial support of the Department of Science and Technology, New Delhi in carrying out this work.

References

[1] M C Cross and P C Hohenberg, *Rev. Mod. Phys.* **65**, 851 (1993)
[2] Y C Lai and R L Winslow, *Physica* **D74**, 353 (1994)
[3] T Shinbrot, *Nonlinear Science Today* **3**, 1 (1993)
[4] H Gang and H Kaifen, *Phys. Rev. Lett.* **71**, 3794 (1993)
[5] E Kostelich, C Grebogi, E Ott and J A Yorke, *Phys. Rev.* **E47**, 305 (1993)
[6] D Auerbach, *Phys. Rev. Lett.* **72**, 1184 (1994)
[7] L Poon and C Grebogi, *Phys. Rev. Lett.* **75**, 4023 (1995)
[8] A Karma, *Phys. Rev. Lett.* **71**, 1103 (1993)
[9] M Bär and M Eiswirth, *Phys. Rev.* **E48**, R1635 (1993)
[10] R Imbihl and G Ertl, *Chem. Rev.* **95**, 697 (1995)
[11] B I Shraiman *et al*, *Physica* **D57**, 241 (1992)
[12] I Aranson, L Aranson, L Kramer and A Weber, *Phys. Rev.* **A46**, 2992 (1992)
[13] I Aranson, H Levine and L Tsimring, *Phys. Rev. Lett.* **72**, 2561 (1994)
[14] J F Lindner, B K Meadows, W L Ditto, M E Inchiosa and A R Bulsara, *Phys. Rev. Lett.* **75**, 3 (1995)
[15] Y Braiman, J F Lindner and W L Ditto, *Nature* (*London*) **378**, 465 (1996)
[16] H D I Abarbanel, R Brown, J J Sidorowich and L Tsimring, *Rev. Mod. Phys.* **65**, 1331 (1993)
[17] H Fujisaka and T Yamada, *Prog. Theor. Phys.* **69**, 32 (1983)
[18] L M Pecora and T L Carroll, *Phys. Rev. Lett.* **64**, 821 (1990)
[19] L M Pecora and T L Carroll, *Phys. Rev.* **A44**, 2374 (1991)
[20] R He and P G Vaidya, *Phys. Rev.* **A46**, 7387 (1992)
[21] K M Cuomo and A V Oppenheim, *Phys. Rev. Lett.* **71**, 65 (1993)
[22] M Ding and E Ott, *Phys. Rev.* **E49**, R945 (1994)
[23] C W Wu and L O Chua, *Int. J. Bifur. Chaos* **4**, 979 (1994)
[24] J F Heagy, T L Carroll and L M Pecora, *Phys. Rev.* **E50**, 1874 (1994)
[25] N F Rulkov, M M Sushchik, L S Tsimring and H D I Abarbanel, *Phys. Rev.* **E51**, 980 (1995)
[26] L Kocarev and U Parlitz, *Phys. Rev. Lett.* **76**, (1996)
[27] T C Newell, P M Alsing, A Gavrielides and V Kovanis, *Phys. Rev. Lett.* **72**, 1647 (1994)
[28] T C Newell, P M Alsing, A Gavrielides and V Kovanis, *Phys. Rev.* **E49**, 313 (1994)
[29] E Ott, C Grebogi and J A Yorke, *Phys. Rev. Lett.* **64**, 1196 (1990)

[30] Y C Lai and C Grebogi, *Phys. Rev.* **E47**, 2357 (1993)
[31] P Grassberger, *Phys. Scr.* **40**, 346 (1989)
[32] G Mayer-Kress and K Kaneko, *J. Stat. Phys.* **54**, 1489 (1989)
[33] A Torcini, A Politi, G Puccioni and G Alessandro, *Physica* **53**, 85 (1991)
[34] H Chate, G Grinstein and L H Tang, *Phys. Rev. Lett.* **74**, 912 (1995)
[35] M Bauer, H Heng and W Martienssen, *Phys. Rev. Lett.* **71**, 521 (1993)
[36] N Parekh, V Ravi Kumar and B D Kulkarni, (submitted)
[37] Y C Lai and C Grebogi, *Phys. Rev.* **E52**, 1894 (1994)
[38] J Warncke, M Bauer and W Martienssen, *Europhys. Lett.* **25**, 323 (1994)
[39] J H Peng, E J Ding, M Ding and W Yang, *Phys. Rev. Lett.* **76**, 904 (1996)
[40] N Parekh, V Ravi Kumar and B D Kulkarni, *Physica* **A224**, 369 (1996)
[41] V Castets, E Dulos, J Boissonade and P De Kepper, *Phys. Rev. Lett.* **64**, 2953 (1990)
[42] Y Kuramoto, *Chemical oscillations, waves and turbulence* (Berlin: Springer) (1984)
[43] G Nicolis, *J. Phys.* **C2**, SA47 (1990)
[44] W Ouyang and H L Swinney, *Nature* (*London*) **352**, 610 (1991)
[45] P Gray and S Scott, *Chem. Eng. Sci.* **38**, 29 (1983)
[46] J E Pearson, *Science* **261**, 189 (1993)
[47] K J Lee, W D McCormick, J E Pearson and H L Swinney, *Nature* (*London*) **369**, 215 (1994)
[48] N Parekh, V Ravi Kumar and B D Kulkarni, *Phys. Rev.* **E52**, 5100 (1995)
[49] D T Lynch, *Chem. Engg. Sci.* **47**, 4435 (1992)
[50] D Horváth, Valery Petrov, S K Scott and K Showalter, *J. Chem. Phys.* **98**, 6332 (1993)
[51] V Petrov, S K Scott and K Showalter, *Philos. Trans. R. Soc. London* **A347**, 631 (1994)
[52] J Argyris, G Faust and M Haase, *An exploration of chaos* (Elsevier Science B V, Amsterdam, 1994)
[53] J L Kaplan and J A Yorke, *Lecture notes in mathematics* **730**, 204 (1979)
[54] S N Rasband, *Chaotic dynamics of nonlinear systems* (Wiley-Interscience, 1989)
[55] Y B Pesin, *Russ. Math. Sur.* **32**, 55 (1977)
[56] H Tong, *Nonlinear time series: A dynamical system approach* (Clarendon Press, Oxford, 1990)
[57] B A Huberman and E Lumer, *IEEE Trans. Circuits Syst.* **37**, 547 (1990)
[58] S Sinha and R Ramaswamy, *Physica* **D43**, 118 (1990)
[59] V Ravi Kumar, B D Kulkarni and P B Deshpande, *Proc. R. Soc. London Ser.* **A433**, 711 (1991)
[60] S Rajashekar and M Lakshmanan, *Int. J. Bifur. Chaos* **2**, 201 (1992)
[61] K Pyragas, *Phys. Lett.* **A181**, 203 (1993)
[62] J K Bandyopadhyay, V Ravi Kumar, B D Kulkarni and P Bhattacharya, *Chem. Engg. Sci.* **48**, 3545 (1993)
[63] H K Qammer, F Mossayebi and L Murphy, *Phys. Lett.* **A178**, 279 (1993)
[64] D Vassiliadis, *Physica* **D71**, 319 (1994)
[65] J K John and R E Amritkar, *Phys. Rev.* **E49**, 4843 (1994)
[66] U Pralitz, *Phys. Rev. Lett.* **76**, 1232 (1996)
[67] H G Bock, *Progress in scientific computing* (Birkhäuser, Boston) 2, 95 (1983)
[68] E Baake, M Baake, H G Bock and K M Briggs, *Phys. Rev.* **A45**, 5524 (1992)
[69] J C Principe, A Rathie and J M Huo, *Int. J. Bifur. Chaos* **2**, 989 (1992)
[70] S A Billings and S Chen, *Neural networks and system identification* (Peter Peregrinus, London, 1992) p. 181

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 1,
January 1997
pp. 325–364

Some recent advances in the theory of homogeneous isotropic turbulence

SUJAN K DHAR, ANIRBAN SAIN, ASHWIN PANDE and RAHUL PANDIT*
Department of Physics, Indian Institute of Science, Bangalore 560 012, India
*Also at Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India

Abstract. We review some advances in the theory of homogeneous, isotropic turbulence. Our emphasis is on the new insights that have been gained from recent numerical studies of the three-dimensional Navier Stokes equation and simpler shell models for turbulence. In particular, we examine the status of multiscaling corrections to Kolmogorov scaling, extended self similarity, generalized extended self similarity, and non-Gaussian probability distributions for velocity differences and related quantities. We recount our recent proposal of a wave-vector-space version of generalized extended self similarity and show how it allows us to explore an intriguing and apparently universal crossover from inertial- to dissipation-range asymptotics.

Keywords. Isotropic turbulence; homogeneous turbulence; turbulence simulation and modelling; theory and models of chaotic systems; critical point phenomena.

PACS Nos 47.27; 05.45; 05.70

1. Introduction

If a fluid is forced through a grid (say in a wind tunnel) or through an orifice, the flow downstream is turbulent [1]. Most of the energy is contained in the large scales of the flow; transport too is dominated by the large scales. Nevertheless, there has been great interest in the study of turbulence at small scales; these include the *dissipation* range, in which energy is dissipated, and the *inertial* range, that lies in between dissipation and large scales. There are both practical and fundamental reasons for this continuing interest. We concentrate on the latter here and refer the reader to the recent article by Sreenivasan and Antonia [2] for a more detailed discussion. Other recent reviews include those by Siggia [3] and Nelkin [4], the brief overview by L'vov and Procaccia [5], and the books by McComb [6], Frisch [7], and Bohr, Jensen, Paladin, and Vulpiani [8]; older material can be found in the classics by Batchelor [9], Tennnekes and Lumley [10], and Monin and Yaglom [11].

There are many fundamental reasons for studying the statistical properties of small-scale turbulence, which is (in many cases, e.g., far from boundaries, etc.), to a good approximation [2, 12, 13], *homogeneous* and *isotropic*: (1) These properties are believed to be *universal*, i.e., independent of the details of the flow; specifically, the asymptotic behaviours of certain correlation functions (the *structure functions* defined below) are universal in a way that is reminiscent of, but perhaps richer than, the universality of correlation functions at continuous phase transitions [14]. (2) From the point of view of dynamical systems, turbulence is the quintessential example of spatiotemporal chaos [8] in an extended, deterministic, driven system. (3) To the extent that the chaos generated by turbulence can be mimicked by an external noise [15], one can try to obtain the properties of correlation functions by studying effective stochastic equations, which are amenable to analytical studies that use renormalization-group, field-theoretical, or closure techniques [16–20]. (4) And last, the scale-invariance of correlation functions in the inertial range is an example of self-organized criticality [21], though very little has been gained by thinking about turbulence in this manner.

This paper is organized as follows: Section 2 gives a brief, elementary overview of the phenomenology of homogeneous, isotropic turbulence to establish notations. Section 3 introduces some of the deterministic models that have been used in theoretical and numerical work and reviews what has emerged from such studies. We make some concluding remarks in § 4.

2. The phenomenology of homogeneous isotropic turbulence

Experiments on turbulent flows often measure the velocity $\mathbf{v}(\mathbf{r}, t)$ or the energy dissipation ϵ at a given spatial point $\mathbf{r}$ at time t, by using probes such as hot-wire anemometers. These are also augmented by flow-visualisation measurements [1]. Therefore, a typical experiment yields a time series for $\mathbf{v}$ or ϵ at a given spatial point. Most experiments are done with a mean background flow with velocity $\mathbf{U}$ and the resulting temporal information is converted into spatial information by using Taylor's *frozen-flow* hypothesis: This states that the turbulence is advected by the mean flow without significant distortion, so a temporal delay Δt can be converted into a spatial separation (along the mean flow) $\Delta x = \Delta t U$ to obtain spatial correlation functions. Such a hypothesis is reasonable when $|\mathbf{U}| \gg v_{\mathrm{rms}}$, the root-mean-square fluctuations in the velocity because of the turbulence; for a more critical evaluation of its validity we refer the reader to refs [2, 4, 7]. The Reynolds number $\mathrm{Re} = UL/\nu$ is the relevant control parameter, which, when large, leads to turbulence; here U and L are typical velocity and length scales, respectively, and ν is the kinematic viscosity. The Taylor-microscale Reynolds number Re_λ, which is useful in comparing widely different systems, will be defined in § 3.

Another important approximation is made in obtaining the energy dissipation $\epsilon \equiv \frac{\nu}{2}\sum_{ij}(\partial_i \mathrm{v}_j + \partial_j \mathrm{v}_i)^2$, where the subscripts i and j denote Cartesian components and ν is the kinematic viscosity: It turns out to be difficult to measure ϵ, so, typically, the *surrogate* $\epsilon' \sim (\partial u_1/\partial t)^2$ is measured, with u_1 the velocity in the streamwise direction; again we recommend [2] for a discussion of the validity of using this approximation.

Given such data, a variety of information can be extracted about the turbulent flow in question. We concentrate on measurements that can be classified into three rough categories:

1. The first is predominantly qualitative information about *coherent structures*, such as vortex streets, eddies, etc. These are helpful, as we will see below, in developing physically appealing scenarios for turbulent cascades, intermittency, and multifractal dissipation.
2. The second is the measurement of spatial correlation functions (typically via the Taylor hypothesis), the most important ones being the order-p velocity *structure*

$$\mathcal{S}_p(r) \equiv \langle [\mathbf{v}_i(\mathbf{x}+\mathbf{r}) - \mathbf{v}_i(\mathbf{x})]^p \rangle. \tag{1}$$

The angular brackets imply, in principle, a spatiotemporal average over the statistical steady state that obtains in the turbulent fluid. In experiments, often only the component parallel to the mean flow is used. Some workers use a modulus instead of the square brackets in (1), but others [2] discourage its use; it clearly makes no difference for even p but can affect exponents [22] for odd p. Other commonly measured correlation functions include the energy spectrum $E(k) \equiv 4\pi k^2 \langle |\mathbf{v}(\mathbf{k})|^2 \rangle$, where $\mathbf{v}(\mathbf{k})$ is the spatial Fourier transform of the velocity, $\mathbf{k}$ the wave vector, and $k \equiv |\mathbf{k}|$.

3. The third is the measurement of probability distribution functions for velocity differences, such as $\delta \mathbf{v}_i \equiv [\mathbf{v}_i(\mathbf{x}+\mathbf{r}) - \mathbf{v}_i(\mathbf{r})]$ (in experiments i is often the direction of the mean flow), the energy dissipation ϵ, etc.

2.1 *Flow visualization*

The first category of measurements, namely, flow visualization, have shown that fluid turbulence is typically associated with *coherent structures*. These can be eddies, as in mixing-layer experiments, or plumes, as in Rayleigh–Béenard systems, etc. [1]. These structures, though clear to the eye, are difficult to quantify and define. In recent elegant experiments [23, 24, 25] high-vorticity, low-pressure filaments, of the type seen in direct numerical simulations (§ 3), have been visualized directly in different turbulent flows.

Many years ago Richardson [26] suggested the following *cascade* picture: Turbulent flow is considered to be made up of a hierarchy of *eddies* (disturbances or inhomogeneities in the flow) of various length scales. Energy is added to the turbulence in the *integral range*, which corresponds to large length scales or small wave numbers, via large-scale processes like stirring. The nonlinear interactions in the fluid cause this energy to flow towards small scales (large wave numbers), by the formation of successively smaller eddies: large eddies lose stability, form smaller eddies, which in turn lose stability, break into even smaller eddies, and so on. In this process, the energy cascades from large to small length scales, with negligible energy dissipation in the inertial range. Viscous processes lead to significant dissipation once the sizes of eddies become comparable to, or smaller than, the dissipation scale $\eta_{\mathrm{d}} \equiv (\nu^3/\langle\epsilon\rangle)^{1/4}$.

2.2 *Structure functions*

If the Reynolds number Re is sufficiently large then there is a substantial inertial range, in which there can be universal scaling behaviour as noted first by Kolmogorov [27] in 1941 (henceforth K41). His theory is based on the following premises:

1. On small scales and far away from boundaries fluctuations in a turbulent flow are homogeneous and isotropic, statistically steady, and universally determined by the mean energy dissipation rate (per unit mass of fluid) ϵ and the kinematic viscosity ν.

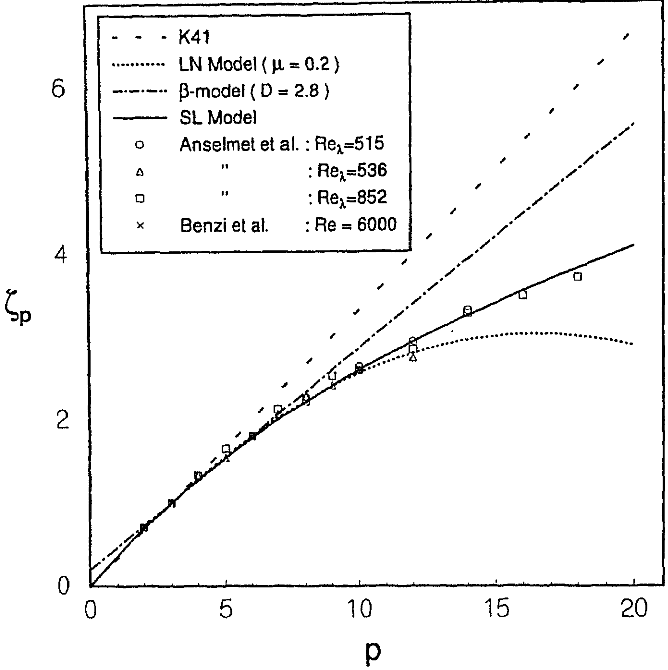


Figure 1. A plot comparing data for the inertial-range exponent ζ_p with different theoretical predictions. Experimental data are from Anselmet *et al* (1984) and Benzi *et al* (1993). Theoretical predictions are from K41, the Kolmogorov log-normal model [42], with $\mu = 0.2$, the β model (with $D = 2.8$), and the SL model (see text). Clearly the SL prediction agrees best with experimental data.

2. In the inertial range, viscosity does not play a significant part, so the statistical properties at a given length scale ℓ are determined by ϵ (which is finite, and non-zero) and the length scale ℓ itself.

Given these assumptions, correlation functions assume simple and universal scaling forms in the inertial range, which follow via dimensional analysis. (Note that the dimensions of ϵ and the energy spectrum $E(k)$, which measures the distribution of energy over different length scales, are, respectively, L^2T^{-3} and L^3T^{-2}.) Specifically

$$\mathcal{S}_p(r) \approx C_p(\langle\epsilon\rangle r)^{p/3}, \tag{2}$$

where C_p are universal amplitudes; the case $p = 3$ yields the Kolmogorov 4/5 law ($C_3 = -4/5$). Also

$$E(k) = C\langle\epsilon\rangle^{2/3}k^{-5/3}, \tag{3}$$

which is the celebrated $-5/3$ law.

The second category of experimental measurements comprise studies of such structure functions. They find that low-order structure functions ($p \leq 3$) are in close agreement with the predictions of K41 [28]. In particular, the $-5/3$ law has been checked [2, 13, 29] over several decades of wavenumber k; the Kolmogorov constant is found to be [2, 30]

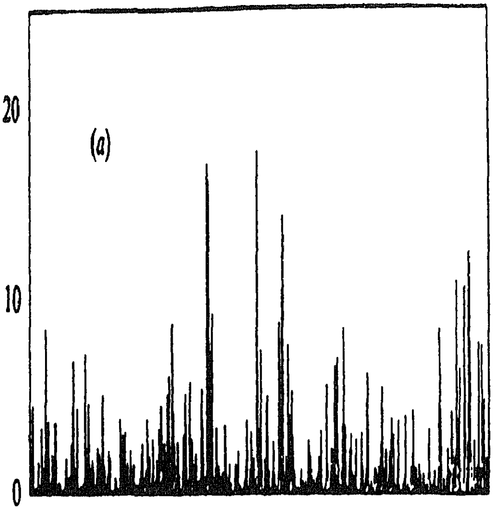


Figure 2. Typical signals of a representative component of the surrogate for the energy dissipation ϵ' (normalized by its mean) versus time t from an atmospheric boundary layer at a moderate Reynolds number (after Meneveau and Sreenivasan [50]).

$C = 0.5 \pm 0.05$. However, the kurtosis $K \equiv \mathcal{S}_4(r)/(\mathcal{S}_2(r))^2$, which is a measure of the *flatness* of the distribution of velocity differences, has values larger than the K41 prediction as $r \rightarrow 0$ [11], which indicates that the velocity gradients are intermittent (i.e., large values of velocity differences are more probable than in K41). This suggests, in turn, that the energy dissipation is nonuniform, contrary to one of the assumptions made by Kolmogorov in K41.

The validity of K41 can be tested experimentally by the measurement of high-order ($p \gg 3$) structure functions, which can capture rare events. Such experiments show that

$$\mathcal{S}_p(r) \sim r^{\zeta_p}; \tag{4}$$

the K41 prediction is $\zeta_p = \zeta_p^{K41} = p/3$. Even though the measurement of high-order structure functions is a difficult task, there are good data from the recent experiments [31, 32, 33] which show that: (1) the ζ_p's are monotonically increasing, nonlinear functions of p (i.e., we have multiscaling); (2) the deviation from the K41 prediction $\delta\zeta_p \equiv p/3 - \zeta_p$ is an increasing function of p and starts becoming significant for $p \geq 4$ (figure 1).

We mention in passing that some experimental studies [34] indicate that $\delta\zeta_p$ decreases with increasing Re, as also suggested by some theories [35, 36]; there are also suggestions [37] of $(\ln \mathrm{Re})^{-1}$ corrections to K41. However, we believe it is fair to say that the general experimental position [2, 33] is that the multiscaling corrections to K41 persist. A careful study over a wide range of Re, possible, e.g., in flows in Helium gas [38, 39], should be able to resolve this potentially significant point. Another intriguing issue has been raised by recent experiments on both open [38] and closed [39] flows in Helium gas: these provide some evidence of a transition or a crossover (in, e.g., the kurtosis K) with increasing Re. Careful studies, already under way [40], must resolve whether these transitions are real or arise because of the limitations of responses of the probes used in these measurements.

How might multiscaling corrections to K41 arise? Experimental data (figure 2) show that the energy dissipation rate ϵ is strongly intermittent. This is ignored in the K41 theory. High-order velocity structure functions are sensitive to the rare but large events which lead to intermittency, hence the discrepancy between K41 and the experimental ζ_p's for large p.

Many phenomenological theories have been proposed to account for the intermittency of ϵ. The oldest is the refined similarity hypothesis (RSH) of Kolmogorov [41], which replaces $\langle\epsilon\rangle$ in (2) by ϵ_r, the energy dissipation rate averaged over a sphere of radius r, i.e.,

$$\mathcal{S}_p(r) \sim \langle \epsilon_r^{p/3} \rangle r^{p/3}. \tag{5}$$

The validity of the RSH has been tested to some extent [42–45], but the issue is certainly not closed as noted by Chen *et al* [46] on the basis of direct numerical simulations (§ 3) and Sreenivasan and Antonia [2]. However, it is often used as a good working hypothesis. If, in the inertial range, we further assume the scaling form

$$\langle \epsilon_r^p \rangle \sim r^{\tau_p}, \tag{6}$$

it follows from (4) and (5) that

$$\zeta_p = p/3 + \tau_{p/3}. \tag{7}$$

Note that the RSH yields a relation between the statistics of velocity differences and the *unknown* behaviour of $\langle \epsilon_r^{p/3} \rangle$. An assumption about the latter yields τ_p and hence ζ_p. In his 1962 theory, Kolmogorov assumed that $\epsilon_r/\langle\epsilon\rangle$ has a log-normal distribution, which, along with the RSH, and the constraint $\tau_1 = 0$, gives

$$\tau_p^{K62} = \mu p(1-p)/2, \tag{8}$$

or

$$\zeta_p^{K62} = \frac{p}{3} + \mu\frac{p}{6}\left(1 - \frac{p}{3}\right), \tag{9}$$

where the intermittency exponent $\mu = -\tau_2 \simeq 0.2$. This yields [47] better agreement with experiments (figure 1) than K41; however, for large p, deviations are noticeable and, as many authors have noted [2], there are several problems with a log-normal distribution for $\epsilon_r/\langle\epsilon\rangle$.

Other phenomenological theories include the β model, proposed by Frisch, Sulem and Nelkin [48], and the multifractal description of turbulence, postulated by Parisi and Frisch [49]. The β model yields

$$\begin{aligned} \zeta_p^{\beta\,\text{model}} &= p/3 + (3-D)(1-p/3), \\ E(k)^{\beta\,\text{model}} &\sim k^{-5/3-(1-D/3)}. \end{aligned} \tag{10}$$

Here $D \equiv \ln\beta/\ln b$, identified with the *fractal dimension* of the space filled by the eddies (at each stage of the cascade the linear size of the eddy is scaled down by a factor $b > 1$ and β is the ratio between the volumes of mother and daughter eddies), is used as a fitting parameter. Figure 1 shows a fit to ζ_p with $D = 2.8$; clearly, the β-model fit overestimates

ζ_p for $8 \lesssim p$. It also yields an inertial-range energy spectrum which is steeper than the $-5/3$ spectrum of K41.

It is easy to see [8] that the Navier–Stokes equation (see below) for an incompressible fluid is invariant under the scale transformation $r \to br$, $v \to b^h v$, $\nu \to b^{1+h}\nu$, $t \to b^{1-h}t$, with an arbitrary exponent h. It has been suggested [49] that, at high Re, i.e., small ν, one must account for multifractality with many exponents h as has been shown explicitly [50] for ϵ (see below): Global scale-invariance (as in K41) is replaced by local scale-invariant structures, i.e., the scaling, or Hölder, exponent h varies at different points in the fluid, which is considered to be a superposition of fractal subsets, each with an h-dependent dimension $D(h)$. Structure functions, which now have contributions from each fractal subset, can be written as

$$\mathcal{S}_p(r) \sim \int (r/L_I)^{hp+3-D(h)} \mathrm{d}h, \tag{11}$$

where L_I is the integral scale. A saddle-point estimate yields (cf., (4))

$$\zeta_p = \min_h [hp + 3 - D(h)] \tag{12}$$

or

$$D(h) - 3 = \max_p [ph - \zeta_p]. \tag{13}$$

The K41 theory corresponds to a single exponent $h = 1/3$ with $D(h) = 3$.

The emergence of multifractal scaling for ϵ through an asymmetric breakdown of eddies has been discussed in detail and substantiated experimentally in the elegant studies of Meneveau and Sreenivasan [50]. In their picture, an eddy of size ℓ in d dimensions breaks up into 2^d eddies of equal size $\ell/2$. The key assumption in this model is that the energy flow to these smaller eddies takes place asymmetrically. The simplest nontrivial choice is that a fraction p_1 of the total energy gets distributed equally to half of these new 2^d eddies; the remaining fraction ($p_2 = 1 - p_1$) of energy goes to the other half. These eddies, in turn, break up into smaller eddies, and this process is repeated with fixed p_1, until one obtains eddies of the order of the dissipation scale. A one-dimensional section of the resulting multifractal set should look qualitatively similar to the energy dissipation signal shown in figure 2. Indeed, Meneveau and Sreenivasan [50] have shown that such a multiplicative process with $p_1 = 0.7$ yields a multifractal set which has the same singularity spectrum as the one obtained by a multifractal analysis of the data in figure 2.

Recently, She and Leveque (SL) [51] proposed a simple model which leads to a prediction for ζ_p in very good agreement with experimental results. This model is based on a set of hypotheses about the statistics of small-scale fluctuations in turbulence. These lead to a formula for ζ_p, which has no adjustable parameters. The SL hypotheses are:

(i) The RSH (5) and (6) is assumed to hold.

(ii) Furthermore the moments of the energy dissipation are assumed to obey the hierarchy

$$\frac{\langle \epsilon_\ell^{p+1} \rangle}{\epsilon_\ell^{(\infty)} \langle \epsilon_\ell^p \rangle} = A_p \left(\frac{\langle \epsilon_\ell^p \rangle}{\epsilon_\ell^{(\infty)} \langle \epsilon_\ell^{p-1} \rangle} \right)^{\beta}, \quad 0 < \beta < 1, \tag{14}$$

where A_p are constants and $\epsilon_\ell^{(\infty)} \equiv \lim_{p\to\infty}\langle\epsilon_\ell^{p+1}\rangle/\langle\epsilon_\ell^p\rangle$ is a quantity that is sensitive to the tail of the distribution of ϵ_ℓ, i.e., to very intermittent structures. SL postulate that the hierarchy originates from some hidden statistical symmetry of the Navier–Stokes equation.

(iii) The quantity $\epsilon_\ell^{(\infty)}$ appearing in the moment hierarchy (14) is associated with the most intermittent dissipative structures and SL suggest that it has a divergent scale dependence, namely,

$$\epsilon_\ell^{(\infty)} \sim \ell^{-\Delta}, \quad \ell \to 0. \tag{15}$$

They argue further that $\epsilon_\ell^{(\infty)} \sim \delta E^\infty/t_\ell$, where $\delta E^\infty \sim \delta v_0^2 \sim \langle\epsilon\rangle^{2/3}\ell_0^{2/3}$ is the largest amount of energy available for dissipation (i.e., the energy in the largest structures) and the scale-dependent energy-transfer time $t_\ell \sim \langle\epsilon\rangle^{-1/3}\ell^{2/3}$. The last asymptotic relation follows from dimensional analysis and the assumption that t_ℓ shows no anomalous scaling. Finally one obtains $\Delta = 2/3$.

(iv) The most intermittent dissipative structures are assumed to be filamentary, i.e., of spatial dimension 1 or codimension 2.

Equations (6) and (14) yield a difference equation for τ_p which, when solved with the boundary conditions that follow from the SL hypotheses, gives

$$\tau_p^{\mathrm{SL}} = -\tfrac{2}{3}p + 2[1 - (\tfrac{2}{3})^p], \tag{16}$$

or, via (7),

$$\zeta_p^{\mathrm{SL}} = \frac{p}{9} + 2[1 - (\tfrac{2}{3})^p]. \tag{17}$$

The most important boundary condition follows from (14) and (15) and is $\lim_{p\to\infty}(\tau_{p+1} - \tau_p) = -2/3$, i.e., $\tau_p \to -2p/3 + C_0$ as $p \to \infty$; it is easy to see from (13) that C_0 is the codimension of the most singular structure, i.e., the one with the smallest Hölder exponent $h_{\min}$. Assumption (iv) yields $C_0 = 2$ which corresponds to $h_{\min} = 1/9$.

The SL formula is in good agreement with the experimental values of ζ_p as can be seen from figure 1. The success of this formula is indeed remarkable, since the validity of the assumptions made is not yet clear. Of the assumptions (i)–(iv), (i) is generally accepted to be a good working approximation. Direct numerical simulations indicate that high-vorticity regions are filamentary [52, 53], which is used to motivate assumption (iv); we return to this point in § 3. Assumption (iii) awaits detailed experimental and numerical verification. Recently, this hierarchy has been verified experimentally [54] to some extent using wind-tunnel data. Dubrulle [55] mentions that a similar hierarchy has been shown to exist in the GOY shell model (see below).

Benzi *et al* [32] have proposed a very interesting way of extending the region over which inertial-range exponents may be extracted from structure functions. The extended self-similarity (ESS), as the procedure is called, states that the real-space structure functions present an *extended* inertial scaling range in log-log plots of one structure function against another. From (4) it follows that

$$\mathcal{S}_p(r) \sim \mathcal{S}_q(r)^{\zeta_p/\zeta_q}; \tag{18}$$

and Benzi *et al* [32] have shown, by using experimental data, that the inertial scaling region extends up to $r \simeq 5\,\eta_d$ in ESS plots. (By contrast, plots of $\mathcal{S}_p(r)$ versus r show inertial range scaling only up to $r \simeq 20\,\eta_d$.) It is conventional to use $q = 3$ in (18), so $\zeta_p/\zeta_3 \simeq \zeta_p$ since ζ_3 is very nearly equal to unity, when the flow is homogeneous and isotropic. (Indeed the Karman–Howarth relation [7] yields $\zeta_3 = 1$.) Thus ESS provides an effective method for determining ζ_p from data for which the extent of the inertial range is often restricted. However, the reason behind the success of ESS is still unclear (but see [56, 57] for some progress on this front).

Although ESS is an effective method for determining inertial-range scaling behaviour in some turbulent flows, it does not work well for all such flows, particularly in the presence of a strong shear. For such systems, Benzi *et al* [58] have proposed the generalized extended self similarity (GESS) ansatz, which they claim extends the scaling down to the smallest resolvable spatial separations. They begin with the dimensionless, real-space structure functions

$$\mathcal{G}_p(r) = \frac{\mathcal{S}_p(r)}{\mathcal{S}_3(r)^{p/3}} \,. \tag{19}$$

If K41 were true, $\mathcal{G}_p(r)$ would be independent of r in the inertial range. However, because of multiscaling, $\mathcal{G}_p(r)$ displays nontrivial scaling behaviour. Specifically, Benzi *et al* [58] propose that $\mathcal{G}_p(r)$ is *self-scaling* over all resolvable length scales, i.e.,

$$\mathcal{G}_p(r) = [\mathcal{G}_q(r)]^{\rho_{p,q}} \,. \tag{20}$$

From the r-dependence of $\mathcal{S}_p$ (4) it follows that

$$\rho_{p,q} = \frac{\zeta_p - p\zeta_3/3}{\zeta_q - q\zeta_3/3} \,. \tag{21}$$

Thus, if $\log \mathcal{G}_p(r)$ is plotted against $\log \mathcal{G}_q(r)$ one should get a straight line with slope $\rho_{p,q}$ up to all resolvable length scales. This has been checked to some extent by Benzi *et al* [58] from numerical and experimental data for $(p, q) = (5, 6)$.

Dissipation-range behaviors of structure functions have been studied less than their counterparts in the inertial range. There are some data for the energy spectrum $E(k)$. In particular, K41 suggests that

$$E(k) = \langle\epsilon\rangle^{2/3} k^{-5/3} f(k/k_{\mathrm{d}}), \tag{22}$$

where the dissipation scale $k_{\mathrm{d}}^{-1} \equiv \eta_{\mathrm{d}}$ and the scaling function $f(k/k_d) \to C$, for small k/k_{d}. The general form of $f(k\eta_{\mathrm{d}})$ is not known, but the consensus from experiments [2,13] and direct numerical studies [59] is that, in the far dissipation range, i.e., for large k/k_{d}, $E(k) \sim k^{\delta} \exp(-ck/k_{\mathrm{d}})$, a form first suggested by Kraichnan [60]. The exponent δ and the constant c are not easy to determine; the best numerical studies [59] yield $\delta \simeq 3.3$ and $c \simeq 7.1$. It is also worth noting that $f(k/k_{\mathrm{d}})$ does not decay monotonically with k/k_{d}; rather, it increases mildly in the inertial range and then starts decreasing in the dissipation range leading to a slight pile up in the energy spectrum in the crossover regime. This has been referred to as a *bottleneck* phenomenon [61].

Soon after the introduction of ESS it was pointed out by Stolovitzky and Sreenivasan [22] that ESS plots cross over from their inertial-range asymptotic behaviour to another asymptotic behaviour in the far dissipation range that is characterized by a steeper power law (cf. (18)). This has not been explored experimentally in the GESS context. We will study this in detail in § 3 in the context of our wave-vector-space (henceforth k space) generalizations of ESS and GESS.

2.3 *Probability distributions*

The third category of measurements yield probability distributions for quantities such as the velocity, components of velocity differences (i.e., $\delta \mathbf{v}_i \equiv [\mathbf{v}_i(\mathbf{x}+\mathbf{r})-\mathbf{v}_i(\mathbf{x})]$), etc.

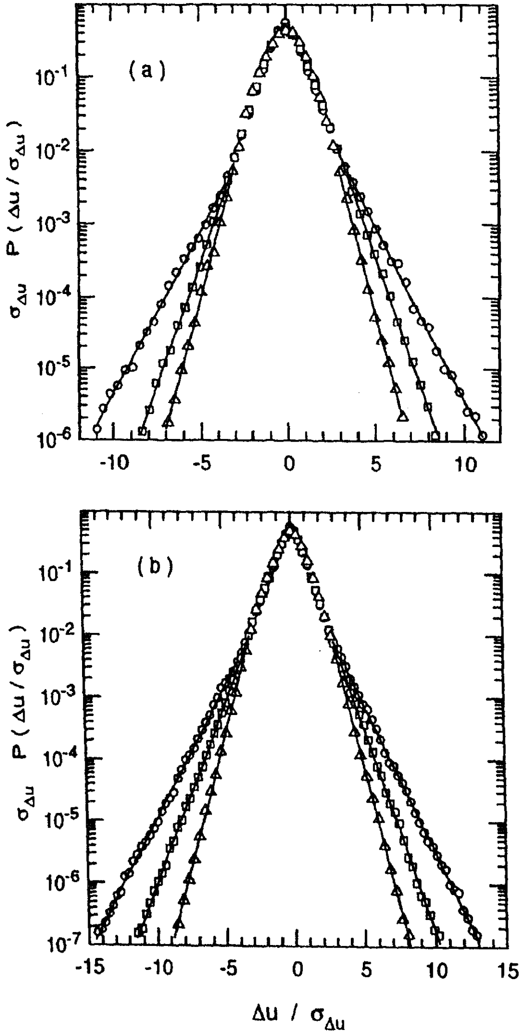


Figure 3. PDFs of velocity differences for separations r within the inertial range for **(a)** wind-tunnel data: $r/\eta_d = 37.5$ (○), 351 (□), and 1870 (△); **(b)** atmospheric surface layer: $r/\eta_d = 61.3$ (○), 408 (□), and 2900 (△). The solid lines correspond to best fits which are of the exponential form (after Praskovsky and Oncley [62]).

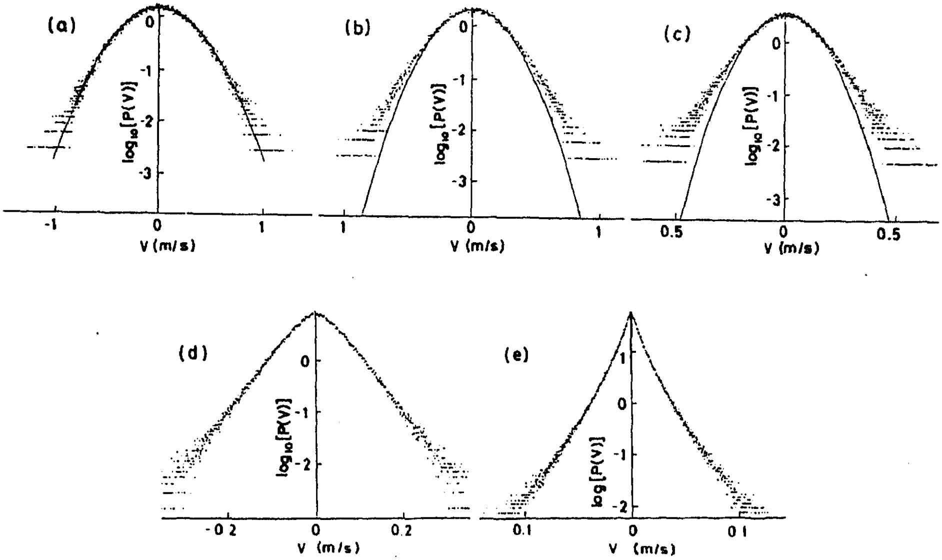


Figure 4. PDFs of band-pass-filtered signals with f_c = (**a**) 0.07 kHz, (**b**) 0.15 kHz, (**c**) 0.70 kHz, (**d**) 3.0 kHz, and (**e**) 10 kHz (see text). Solid curves show Gaussian distributions (after Katsuyama, Horiuchi and Nagata [34]).

There is considerable evidence [62] that the distribution of velocity differences is not Gaussian at small length scales in fully developed turbulence: Experimental data show that the probability distribution functions (PDFs) of velocity gradients over small distances (i.e., large wavenumbers) have a tail which falls more slowly than that of a Gaussian distribution [62, 63]. It is often said [64] that such non-Gaussian behaviour is associated with the presence of small-scale coherent structures which lead to intermittency. Whereas this might well be true, it is worth noting [8] that even K41 yields a non-Gaussian, stretched-exponential distribution for the distribution of velocity gradients; however, the stretching predicted by K41 is less than that observed experimentally. Figure 3 shows a plot of the PDFs of velocity differences obtained by Praskovsky and Oncley [62] from wind-tunnel experiments and from the atmospheric surface layer; they fit well to an exponential form. However, a general theoretical understanding of the precise relation between small-scale coherent structures and such PDFs is still lacking.

Katsuyama, Horiuchi and Nagata [34] have studied the behaviour of PDFs obtained from a band-pass-filtered velocity signal. In this study, the velocity time series, recorded in wind-tunnel experiments, was band-pass filtered to obtain the record $V(t; f_c)$, where f_c is the midband frequency of the filter. The PDFs of $V(t, f_c)$ are Gaussian at small f_c, but decay more slowly than Gaussian distributions as f_c is increased (figure 4). A number of authors had suggested earlier [63, 65] that such PDFs can be approximated by stretched exponentials of the form $P(V) \propto \exp(-\alpha|V|^m)$. For their band-pass-filtered signals, Katsuyama *et al* [34] measured the stretching exponent m and found that, for small f_c, corresponding to integral-range time scales, $m \simeq 2$, i.e., a Gaussian PDF; as f_c was increased (times corresponding to inertial-range timescales), m started falling below 2.

This fall became more and more significant for high f_c's, corresponding to dissipation-range timescales.

Katsuyama *et al* used another method for characterising the deviation of the PDF $V(t; f_c)$ from a Gaussian PDF: They calculated the normalized moments

$$\gamma_{2p} = \frac{(2p-1)!!}{2} \frac{\langle [V(t; f_c)]^2 \rangle^p}{\langle [V(t; f_c)]^{2p} \rangle}, \quad p = 2, 3, 4, \ldots \tag{23}$$

where $(2p-1)!! \equiv (2p-1)(2p-3)\cdots 3.1$. For a Gaussian distribution, $\gamma_{2p} = 0.5$ for all p. Even a very weak deviation from a Gaussian distribution becomes apparent if we calculate γ_{2p} for sufficiently large p. The band-pass-filtered data show that γ_{2p} goes slightly below 0.5 for f_c's corresponding to inertial-range timescales. This is followed by a more rapid decay in dissipation-range timescales and, as anticipated, the deviation from Gaussian behaviour is clearly more prominent for large p.

Recall that, in his 1962 theory [42], Kolmogorov assumed that $\epsilon_r / \langle \epsilon \rangle$ has a log-normal distribution. This is not substantiated by experiments, which yield data that can be fit [50] to a stretched-exponential form with a stretching exponent of $1/2$; however, this latter form violates the Carleman criterion, i.e., such a PDF is not determined uniquely by its moments [2, 66]. It has also been shown recently [55, 67] that the SL assumptions lead to distributions in the log-Poisson class. We are not aware of a detailed experimental verification of this.

3. Deterministic models for the study of homogeneous isotropic turbulence

The tasks before a theory of homogeneous, isotropic turbulence are many. Given the perspective adopted in this paper, these must include an understanding of the multiscaling behaviours of the structure functions and probability distributions defined in § 2. It is also important to elucidate the role, if any, of coherent structures in determining these properties. At the moment we are very far from this goal. However, progress has been made over the past decade by using: (a) approximate analytical methods to study stochastic models, which are discussed in the companion paper by Bhattacharjee [20]; and (b) direct numerical studies of deterministic models. We concentrate on the latter in this section, which is organised as follows: It begins with an introduction to direct numerical solutions of the Navier–Stokes equation and the GOY shell model, which are used in studies of homogeneous, isotropic turbulence. This is followed by a brief overview of what has been learnt about intermittent structures, the scaling of structure functions (including ESS and GESS), and various PDFs from such studies. Our emphasis is on showing how studies of both the Navier–Stokes equation and reduced models like the GOY shell model can augment each other as we have shown in recent studies [89] of k-space ESS and GESS.

3.1 *Models*

The model of choice for the study of fluid turbulence is, of course, the deterministically forced Navier–Stokes (NS) equation for an incompressible fluid. If we restrict ourselves to low-Mach-number flows and cases where the equation for the energy density can be

Table 1. Parameters ν (viscosity), ν_{H} (hyperviscosity), Re_λ (Taylor-microscale Reynolds number), τ_{e} (box-size eddy-turnover time), τ_{av} (averaging time), τ_{t} (transient time) and k_{d} (dissipation-scale wavenumber) for our $3d$ NS runs NS1-4 ($L = \pi$, $k_{\max} = 64$) and GOY-model runs G1-8 ($k_{\max} = 2^{22}k_0$). The step size (δt) used is 0.02 for NS1-4, 10^{-4} for G1-4, and 2.10^{-5} for G5-8. In our NS runs we estimate $k_{\mathrm{d}} = \eta_{\mathrm{d}}^{-1}$ by solving the nonlinear equation, obtained by balancing dissipative and convective time scales, $\nu\eta_{\mathrm{d}}^{-2} + \nu_h\eta_{\mathrm{d}}^{-4} = \langle\epsilon\rangle^{1/3}\eta_{\mathrm{d}}^{-2/3}$, which reduces to the formula given above for η_{d} if the hyperviscosity $\nu_h = 0$.

Run	ν	ν_{H}	Re_λ	$\tau_{\mathrm{e}}/\delta t$	$\tau_{\mathrm{t}}/\tau_{\mathrm{e}}$	$\tau_{\mathrm{av}}/\tau_{\mathrm{e}}$	$k_{\max}/k_{\mathrm{d}}$
NS1	5.10^{-4}	0	$\simeq 3.5$	$\simeq 3.10^4$	$\simeq 1$	2	$\simeq 4$
NS2	2.10^{-4}	0	$\simeq 8$	$\simeq 3.10^4$	$\simeq 1$	$\simeq 2.5$	$\simeq 2.3$
NS3	5.10^{-4}	5.10^{-6}	$\simeq 3.5$	$\simeq 3.10^4$	$\simeq 1$	$\simeq 1$	$\simeq 6.5$
NS4	5.10^{-4}	10^{-6}	$\simeq 22$	$\simeq 3.10^3$	$\simeq 10$	$\simeq 7$	$\simeq 2$
G1-4	$5.10^{-6} - 10^{-7}$	0	$4.10^4 - 3.10^5$	$\simeq (1.5 - 2.0)10^4$	$\simeq 500$	$\simeq 2500$	$\simeq 2^5 - 2^3$
G5-8	$5.10^{-8} - 10^{-9}$	0	$3.5.10^5 - 3.10^6$	$\simeq (0.7 - 1)10^5$	$\simeq 500$	$\simeq 2500$	$\simeq 2^3 - 1$

neglected, the NS equation is

$$\partial_t \mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} = \nu\nabla^2\mathbf{v} - \nabla P/\rho + \mathbf{f}\,, \tag{24}$$

where $\mathbf{v}$ is the velocity field, P the pressure, ρ the density, ν the kinematic viscosity, $\mathbf{f}$ the external, deterministic force, and incompressibility is enforced by the condition

$$\nabla\cdot\mathbf{v} = 0. \tag{25}$$

Equations (24) and (25) must be solved with suitable boundary conditions and initial conditions; periodic boundary conditions are convenient in studies of homogeneous, isotropic turbulence. At high Re, little can be said analytically about the statistical steady state of this driven equation beyond the phenomenological theories described in § 2 or by using stochastic models [20]; one must, perforce, obtain information about this steady state from direct numerical solutions (DNS). In studying the properties of homogeneous and isotropic turbulence, it is sometimes convenient to work in k space; here the equation is,

$$\partial_t v_i(\mathbf{p}) + iM_{ijk}(\mathbf{p})\sum_{\mathbf{q}} v_j(\mathbf{q})v_k(\mathbf{p}-\mathbf{q}) = -\nu p^2 v_i(\mathbf{p}) + f_i(\mathbf{p}), \tag{26}$$

with $\mathbf{p}$ and $\mathbf{q}$ denoting wave vectors, $M_{ijk}(\mathbf{p}) \equiv [p_j P_{ik}(\mathbf{p}) + p_k P_{ij}(\mathbf{p})]/2$, $P_{ij}(\mathbf{p}) \equiv [\delta_{ij} - p_i p_j/p^2]$ the transverse projector, and Fourier transforms are implied by wave-vector arguments. Typically one uses periodic boundary conditions and, if a hyperviscosity ν_H (table 1) is used, νp^2 in (26) is replaced by $(\nu + \nu_H p^2)p^2$. In numerical studies it is advantageous to use a pseudo-spectral scheme that evaluates certain terms in real space and others in k space. Such studies began with the pioneering work of Orszag and Patterson [68]. Subsequent studies, carried out by many groups, include refs. [28, 52, 53, 59, 69–76]. We follow Meneguzzi and Vincent [73] for our NS simulation.

In spite of rapid advances in computing power, direct numerical studies of the NS equation remain daunting at high Re. The largest studies have used 512^3 Fourier modes for the velocity field and achieved a Taylor-microscale Reynolds number

$\mathrm{Re}_\lambda \equiv v_{\mathrm{rms}}\lambda/\nu \simeq 220$; here the Taylor-microscale is [73] $\lambda \equiv [\int E(k)\mathrm{d}k/\int k^2 E(k)\mathrm{d}k]^{1/2}$ and $v_{\mathrm{rms}} \equiv [2\int_0^\infty E(k)\mathrm{d}k]^{1/2}$ is the root-mean-square velocity. Larger studies are not feasible yet because computational costs scale as $\mathrm{Re}^{9/4}$, for storage requirements, and $\mathrm{Re}^{1/2}$ for the number of arithmetic operations. Important time scales include the integral and box-size eddy-turnover times defined, respectively, as $\tau_I \equiv L_I/v_{\mathrm{rms}}$ and $\tau_e \equiv L/v_{\mathrm{rms}}$, where $L_I \equiv [\int k^{-1}E(k)\mathrm{d}k/\int E(k)\mathrm{d}k]^{1/2}$ is the integral scale and L is the box size . For our four runs for the NS equation (table 1) $\tau_e \simeq 8\tau_I$; we report all times in terms of the former.

Given these computational constraints it is useful to work with simplified models that are physically motivated and capture as many aspects of fluid turbulence as possible. Though such models must, perforce, make several approximations (see below), they can allow us to explore the large-Re regime, which is crucial for the resolution of many questions about the asymptotic behaviours of structure functions, etc. The most successful models of this type are *shell models*. Of these the one due to Gledzer [77] and Ohkitani and Yamada [78], infelicitously called the GOY model, has been studied most over the past five years; and, as we show below, the GOY model yields, in spite of its simplicity, results that are not only physically instructive but also in remarkably good agreement with experiments.

The Richardson-cascade picture suggests that the scaling behaviour seen in fluid turbulence might well arise in a simplified dynamical system with a reduced number of degrees of freedom arranged hierarchically. This key idea motivated the seminal studies of Obukhov [79], Gledzer [77], and Desnyansky and Novikov [80], who postulated shell models quite similar to, though not as succesful as, the GOY model, on which we concentrate here. Other more complicated shell models are mentioned briefly below; we refer the reader to ref. [8] for a more extensive coverage of these.

Such shell models must have the symmetries of the NS equation. Unfortunately, it is not clear whether we know all the relevant symmetries; but one can at least ensure the essential conservation laws (of energy and helicity) hold in the inviscid and unforced limit. Furthermore, to attain much larger Reynolds numbers than is possible in direct numerical simulations of the NS equation, the "shells" in k space are logarithmically spaced. In the GOY model they are the one-dimensional set of wave vectors k_n, labelled by shell numbers n:

$$k_n = k_0\,\lambda^n, \quad n = 1, 2, \ldots, N, \tag{27}$$

where λ is the ratio between wave vectors in nearest-neighbour shells, k_0 sets the scale of wave vectors, and N is the total number of shells. The dynamical variables are the complex scalar velocities v_n associated with the shells; these can be considered to represent the Fourier transform of the velocity characterising an eddy of scale $\sim k_n^{-1}$. The evolution equation for v_n is

$$\frac{\mathrm{d}}{\mathrm{d}t}v_n + \nu k_n^2 v_n = F_n + iC_n, \tag{28}$$

where ν is the kinematic viscosity, F_n is a forcing term, and C_n is the nonlinear coupling between different shells. Thus the NS equation is approximated by a dynamical system comprising $2N$ ordinary differential equations (each v_n has both real and imaginary parts). Rigid boundary conditions are imposed on the v_n's, i.e., $v_n = 0$ for $n < 1$ and

$n > N$. The forcing term has a constant amplitude and is typically chosen to act on a particular shell n', i.e.,

$$F_n = F_0 \delta_{n,n'}, \tag{29}$$

with n' the index of the shell on which the force acts. As described in earlier studies [78, 81–85], the usual choice is $n' = 4$, but Kadanoff, Lohse and Schörghofer [86] note that the choice $n' = 1$ (which we adopt in our dynamical studies below) does not change the behaviour of the model in the static case. To mimic the advection terms in the NS equation, the nonlinear term C_n in (28) should have terms like $k_n v_{n'} v_{n''}$; but, since the Richardson-cascade picture suggests that an eddy of a particular length scale interacts significantly only with eddies of nearby scales, such interaction terms are chosen to be local in k space (i.e., n' and n'' are chosen to be close to n). The GOY form of C_n is

$$C_n = a_n k_n v^*_{n+1} v^*_{n+2} + b_n k_{n-1} v^*_{n-1} v^*_{n+1} + c_n k_{n-2} v^*_{n-1} v^*_{n-2}, \tag{30}$$

with $a_{N-1} = a_N = b_1 = b_N = c_1 = c_2 = 0$; thus the velocity of a shell is affected only by the velocities of its nearest- and next-nearest-neighbour shells. To specify C_n completely we must fix the constants a_n, b_n, c_n. The first constraint is that the kinetic energy $E = \sum_n |v_n|^2/2$ must be conserved in the inviscid ($\nu = 0$) and unforced ($F_n = 0$ for all n) limit. This yields, $a_n + b_{n+1} + c_{n+2} = 0$. These coefficients are defined modulo a multiplicative factor, so that one can fix $a_n = 1$ without loss of generality, whence

$$a_n = 1; \quad b_n = -\delta; \quad c_n = -1 + \delta. \tag{31}$$

The parameter δ does not have an analogue in NS dynamics, but it plays an important role in the GOY shell model since it determines the ratio between the backward (towards smaller k) and forward (towards larger k) energy cascades as shown by Biferale, Lambert, Lima, and Paladin [87]. The choice $\lambda = (1 - \delta)^{-1}$ (in most studies $\delta = 1/2$ and $\lambda = 2$), in addition to conserving the kinetic energy, also conserves $H = \sum (-)^n k_n |v_n|^2$ in the inviscid and unforced limit [88]. Kadanoff, Lohse, Wang and Benzi [84] have noted that H is the GOY-model analogue of the *helicity*, which is conserved in the inviscid, unforced NS equation. Note also that the term C_n in the GOY model conserves the phase-space volume, i.e., the total volume in the $2N$-dimensional velocity space, since $\partial C_n / \partial v_n = 0$. And, as in the NS equation, the GOY model is forced near small k, but the dissipation occurs principally at large k's, which, along with the nonlinear coupling, ensures a cascade of energy from small to large wavenumbers. Of course the GOY shell model is not derivable from the NS equation in any obvious way. But its simplicity and numerical tractability have inspired quite a few studies, principally numerical simulations, some of which we discuss below. Some limitations of the GOY shell model are discussed towards the end of this section.

It is easy to see that the Kolmogorov scaling relation $v_n \sim k_n^{-1/3}$ for the GOY model is a solution of the static problem in the inviscid and unforced limit. It has been shown [87] that *static* solutions of the GOY model, in the inviscid, unforced limit, are the fixed points of the map $q_n = \delta/2 + (1 - \delta)/4q_{n-1}$, where $q_n \equiv v_{n+3}/v_n$. The K41-type solution, $v_n \sim k_n^{-1/3}$, corresponds to the fixed point $q = 1/2$. For the dynamic but unforced and inviscid case, this K41 fixed point becomes unstable (for $\delta = 1/2$, see below), but continues to play a crucial role for the dynamics with finite viscosity and forcing. Before embarking on a dynamical study of the GOY model, it is, therefore, useful to understand the nature of its static solutions. We refer the reader to refs [87,92] for detailed discussions of such solutions.

Though the GOY-model equations look simple, they cannot be solved analytically. Thus one must resort to direct numerical studies, which must be carried out carefully since the equations form a stiff system. We follow the numerical scheme of ref. [83].

In addition to the choice of the constants a, b, c and λ given in eq. (31), we must specify the forcing term (29). We choose: $F_0 = 5(1+i)10^{-3}$, $n' = 1$, i.e., we force the first shell, not the fourth as in most earlier studies [81–84]. However, we have checked explicitly that the scaling behaviour of the model is the same with $n' = 4$ and $n' = 1$. The advantage of using $n' = 1$ is that the inertial-range is broader, so we obtain better estimates of the inertial-range exponents. Furthermore, we set $k_0 = 2^{-4}$ and work with $N = 22$. Since we set $\lambda = 2$, this leads to a wavenumber range of $2^{-3} \leq k_n \leq 2^{18}$. Earlier studies [82–84] have shown that the choice $N = 22$ leads to a broad wavenumber regime, which, depending on Re, yields sizeable inertial and dissipation ranges. Data for our GOY-model runs are given in table 1.

Biferale and Kerr [90] have pointed out that the GOY shell model has an apparent asymmetry between the odd and even shells. This does not have any counterpart in NS dynamics. To overcome this they have proposed two velocity fields per shell that transport positive- and negative-helicity components, respectively. Studies of such models are in their infancy and we refer the reader to ref. [90] for details. Even more realistic are the hierarchical shell models studied by Eggers and Grossmann [95] and Grossmann and Lohse [96]. Their scheme, also referred to as the reduced wavevector approximation (REWA), will be described briefly below.

We are now in a position to discuss some of the results that have been obtained by direct numerical studies of the NS equation and the GOY model. We begin with visualisations of intermittent structures, then discuss structure functions, ESS, and GESS, and end with probability distributions.

3.2 *The visualisation of intermittent structures*

Siggia [52] was the first to provide numerical evidence that, at large Re, regions of large $|\omega|$ (where the vorticity $\omega \equiv \nabla \times \mathbf{v}$) are fairly long-lived, tube-like structures. The existence of these filamentary, high-vorticity tubes has been confirmed by subsequent high-resolution runs [53, 69, 71, 73]. As the value of $|\omega|$ in an iso-$|\omega|$ plot is decreased, the vorticity tubes thicken until they eventually become sheet-like. This is illustrated in figure 5, which has been obtained from our relatively low-resolution pseudo-spectral simulation that uses 64^3 Fourier modes. In figure 5 we superimpose the velocity field (white arrows) on the iso-$|\omega|$ surfaces and in their vicinities; the circulation of the velocity vectors is as expected around regions of high vorticity. She *et al* [53] have noted that the velocity field around such vortex tubes exhibits swirling motion with a sizeable component along the local vortex-tube axis, which might lead to a local weakening of nonlinear effects [2, 93]. The viewing angles in these photographs have been chosen to display the tubular nature of the filaments clearly.

The existence of such filamentary structures is a crucial input into the She–Leveque model (§ 2). Since this model relies on a set of assumptions for the small-scale behaviour of the energy dissipation ϵ, it is interesting to check whether the existence of vorticity

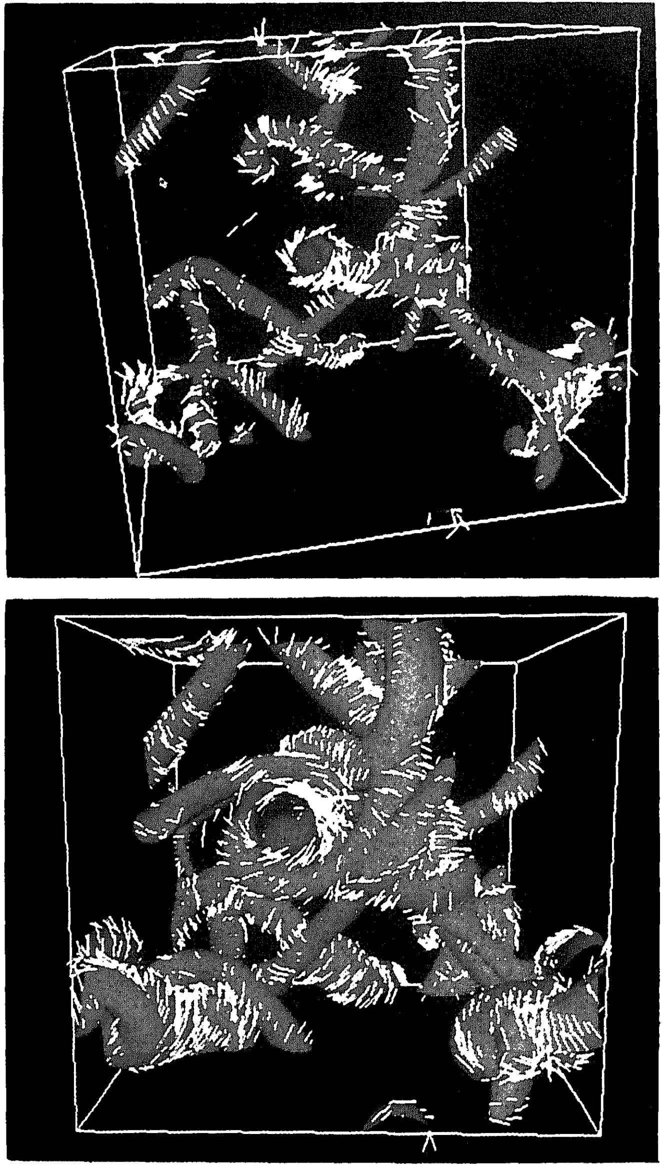


Figure 5. Iso-$|\omega|$ surfaces with superimposed velocity vectors (at points near these surfaces) obtained from an instantaneous snapshot of the velocity and vorticity fields from our run NS4 (table 1). The iso-$|\omega|$ surfaces go through points with fixed $|\omega|/|\omega|_{\max} = \mathcal{O}$, where $|\omega|_{\max}$ is the maximum value of $|\omega|$; the velocity vectors are shown by white arrows whose lengths are proportional to their magnitudes and whose heads are too small to be seen. (The small red arrowhead is a pointer that should be ignored here and in all subsequent photographs.) The red surface (top plate) goes through points with $\mathcal{O} \simeq 0.31$ and the orange surface (bottom plate) goes through points with $\mathcal{O} \simeq 0.12$; this illustrates the thickening of the vorticity filaments with decreasing $|\omega|$. These are perspective views with the simulation box shown by a white boundary; we use periodic boundary conditions.

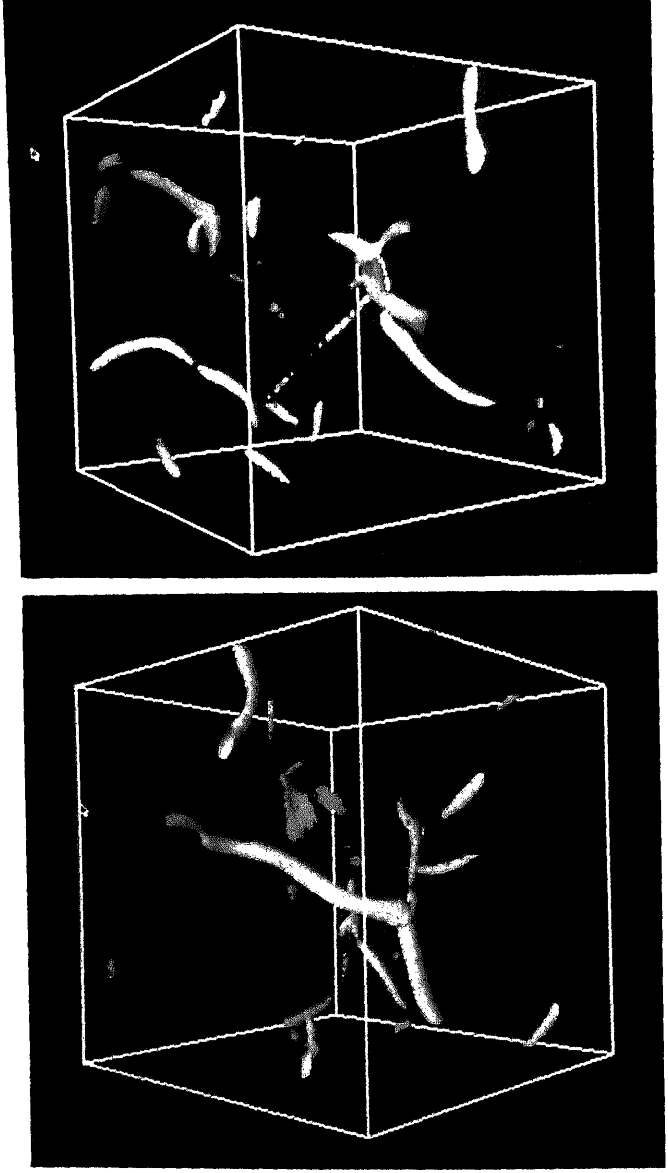


Figure 6. Superimposed iso-$|\omega|$ and iso-ϵ surfaces obtained from an instantaneous snapshot of the vorticity and dissipation fields from our run NS4 (table 1). The cream-coloured surfaces go through points with fixed $|\omega|/|\omega|_{\max} = \mathcal{O}$ whereas blue-green surfaces go through points with fixed $\epsilon/\epsilon_{\max} = \mathcal{E}$ (the subscript max denotes maximum value). Here $\mathcal{O} = \mathcal{E} = 0.5$. The top figure (**a**) is a rotated version of the bottom one (**b**). The iso-$|\omega|$ surfaces clearly show that high-vorticity regions are filamentary; however, the iso-ϵ surfaces seem more like shredded sheets or blobs. Also, high-vorticity and high-dissipation regions do not coincide. These are perspective views with the simulation box shown by a white boundary; we use periodic boundary conditions.

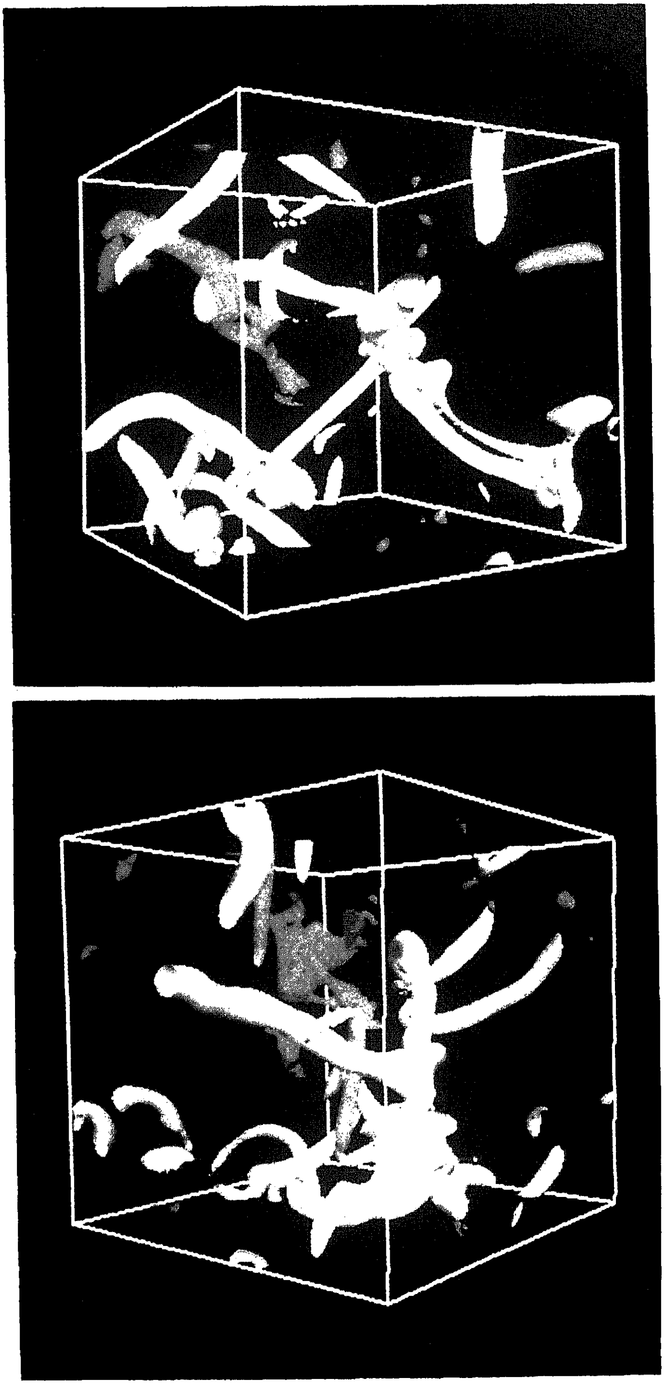


Figure 7. The same as figure 6 but with $\mathcal{O} = \mathcal{E} \simeq 0.33$, so the vorticity tubes are thicker; however, the iso-ϵ surfaces still seem like shredded sheets or blobs. The top figure (**a**) is a rotated version of the bottom one (**b**).

obtained from an instantaneous snapshot of the real-space velocity field in our numerical study of the NS equation (run NS4 in table 1); the ω and ϵ fields follow from this velocity field. In figures 6 and 7 the cream-coloured iso-$|\omega|$ surfaces show well-formed tubes for the $|\omega|$ values used. The blue-green iso-dissipation surfaces show clearly that regions of large ϵ are, by contrast, like small shredded sheets or blobs (cf. [52]), which have no overlap with the high-vorticity tubes. Nor does their codimension seem to be the same as that of the vortex filaments. (Cream-coloured surfaces go through points with fixed $|\omega|/|\omega|_{\max} = \mathcal{O}$ whereas blue-green surfaces go through points with fixed $\epsilon/\epsilon_{\max} = \mathcal{E}$; and, to make the comparison of these surfaces as meaningful as possible, we use equal ratios, i.e., $\mathcal{O} = \mathcal{E}$.) The precise implications of this for the derivation of the SL formula (§ 2) are unclear. If its success in predicting ζ_p is any clue, there might be a way of deriving it without using the precise codimensions of high-dissipation, as opposed to high-vorticity, regions.

It is debatable whether any real-space structures can be obtained reliably from the GOY model, which is defined on a logarithmically discretised k space. Some authors have recently proposed [8] a procedure for generating an artificial velocity field $\mathbf{v}(\mathbf{r}, t)$ in three-dimensional real space from the shell-model velocity fields. The first step is to introduce a three-dimensional wavevector $\mathbf{k} = k_n \mathbf{e}_n$, where k_n is the wavenumber of the n-th shell and the $\mathbf{e}_n$'s vectors of unit norm. Next the j-th component ($j = x, y, z$) of the real-space velocity field is constructed through some kind of an "inverse Fourier transform"

$$v_j(\mathbf{r}, t) = \sum_{n=1}^{N} C_n^{(j)} [v_n(t) e^{i\mathbf{k}_n \cdot \mathbf{r}} + \text{c.c.}], \tag{32}$$

where the coefficients $C_n^{(j)}$ are random numbers of $O(1)$ and c.c. denotes complex conjugate. It can be checked that the resulting v_j's obey the incompressibility condition (25), since the coefficients satisfy $\sum_{j=1}^{3} C_n^{(j)} e_n^{(j)} = 0$, for all n. However, it remains to be checked explicitly whether the velocity fields calculated by using such a prescription conform with those seen experimentally.

3.3 *Structure functions, ESS and GESS*

The scaling of real-space structure functions $S_p(r)$ has been studied via numerical simulations of the NS equation by several groups [71, 73, 76]. Agreement with K41 is good for $p \leq 3$ and $E(k)$ follows the $-5/3$ law well [73]. Some numerical studies have found that the RSH (5) holds reasonably well. However, the most recent study has cautioned that the result depends on which surrogate (see § 1) of ϵ is used. Specifically, if one defines $\epsilon^{ij} \equiv \nu(\partial \mathrm{v}_i/\partial \mathrm{x}_j)^2$, then the RSH holds better with the surrogate ϵ^{11} than with ϵ^{21}, in agreement with recent experiments.

As regards multiscaling, data for ζ_p from direct numerical studies of the NS equation [71, 73] show significant deviations from K41 for $p > 3$; however, it is still not feasible computationally to attain very large Re_λ (the largest so far [76] seems to be $\mathrm{Re}_\lambda = 216$). Unfortunately this is not large enough to settle certain issues of principle: For example, does $\delta\zeta_p = p/3 - \zeta_p \to 0$ as $\mathrm{Re} \to \infty$, as suggested by some studies [34–36]? This is best done by using simpler models like the GOY shell model described above; of course, one must then contend with the issue of the relation of shell-model and NS results, which we discuss later.

The velocity structure functions for the GOY shell model are

$$S_p(k_n) \equiv \langle |v_n|^p \rangle. \tag{33}$$

Since the k_n's in the GOY shell model are logarithmically spaced, the GOY-model analogue of the energy spectrum is $E(k) = |v_n|^2/k_n$ and the exponents ζ_p are given by

$$S_p(k_n) \sim k_n^{-\zeta_p}. \tag{34}$$

It is natural in the GOY model to use these k-space structure functions. Since we want to use direct numerical simulations of the NS equation and the GOY model in complementary ways, it is useful to study k-space structure functions in the NS context. We define these as

$$S_p(k) \equiv \langle |\mathbf{v}(\mathbf{k})|^p \rangle. \tag{35}$$

We use the convention that calligraphic symbols like $\mathcal{S}_p$ denote real-space structure functions and Roman symbols like S_p their k-space *analogues* (not Fourier transforms). In a recent study we have provided numerical evidence that yields [89]

$$S_p(k) \sim k^{-(\zeta_p + 3p/2)}, \tag{36}$$

for k in the inertial range. This difference between GOY-model and NS k-space structure functions arises because of phase space reasons (i.e., the logarithmic discretisation of k space in the former and the usual discretisation in the latter). To have a meaningful comparison between these two models, we also define GOY-model analogues of the Taylor-microscale and the root-mean-square velocity (see above for the $3d$ NS equation): $\lambda = 2\pi/k_1[(\sum_n |v_n|^2)/(\sum_n k_n^2|v_n|^2)]^{1/2}$ and $v_{\mathrm{rms}} = [(2\sum_n |v_n|^2)/(2\pi/k_1)]^{1/2}$, which can be used to obtain the Taylor-microscale Reynolds number $\mathrm{Re}_\lambda = v_{\mathrm{rms}}\lambda/\nu$; furthermore, we use (cf., [83]) $E(k) = S_{n,2}/k_n$. The time-scale associated with the smallest wavenumber is $\sim k_1^{-1}/|v_1|$; when averaged over the steady state it gives the large-scale eddy turnover time τ_e, which is the analogue of the box-size eddy-turnover time in the NS case. For our GOY-model runs we find $\mathrm{Re}_\lambda \sim \nu^{-0.5}$, as expected [102] at large ν^{-1}. We will return to a comparison of these two models when we discuss our k-space versions of ESS and GESS [89] in inertial and dissipation ranges below.

Our GOY-model runs span the range $10^4 \lesssim \mathrm{Re}_\lambda \lesssim 3 \times 10^6$. In figure 8 we show the energy spectrum as a function of the shell number for our runs G1-8. (Care must be taken to ensure that the coputational runs are sufficiently long to obtain a statistical steady state; see table 1 for our run parameters.) Note that $E(k)$ is plotted on a $\log_{10}$ scale and the shell number is proportional to the logarithm (base 2) of the wavenumber. The slight rise in the curves near $n = 1$ is a vestige of the forcing in the first shell. The straight-line portion of

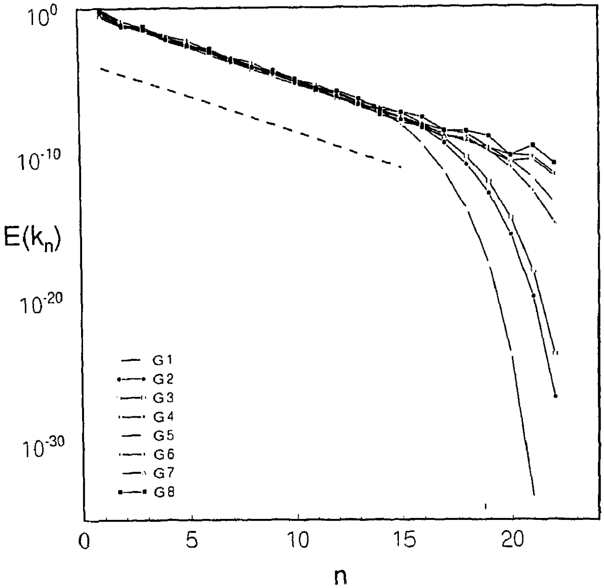


Figure 8. The energy spectrum $E(k_n)$(plotted on a log base 10 scale) as a function of shell number (n), for the runs G1-8 (table 1). The dashed line represents the $k^{-5/3}$ form. The curves for large values of ν (runs G1–G3) start veering off from the $-5/3$ form near the onset of the dissipation range. For small values of ν (runs G6-8), the dissipation range starts beyond $n = 22$.

the plot for intermediate k marks the inertial range and fits the expected $k^{-5/3}$ form well. For runs G1-3, which have the highest viscosities we use, $E(k)$ starts falling from the $k^{-5/3}$ line in figure 8. This marks the beginning of the dissipation range. Note that, for the small-viscosity runs (G5-8), only the beginning of this crossover is visible given our wavenumber range. We will return to the form of $E(k)$ in the dissipation range below.

Many authors [82–84] have used such studies to show that GOY-model structure functions exhibit significant deviations from K41 for $p > 3$ and, for the conventional choice of parameters which conserve both the energy and the helicity in the inviscid, unforced limit, yield exponents ζ_p in good agreement with experiments and the SL formula. (An approximate analytical understanding of multifractality in the GOY shell model may be obtained by a closure scheme as shown by Benzi, Biferale and Parisi [91].) These studies have also pointed out that log-log plots of $S_p(k_n)$ versus k_n show an oscillatory behaviour that is a GOY-model artifact related to an underlying 3-cycle in the static solution [92]. In figure 9 we show such a plot for different values of p from our run G1. The increase in S_p near $n = 1$ arises because we force the first shell. In the region $4 \leq n \leq 14$ the plots are nearly linear, suggesting the algebraic decay of $S_p(k_n)$ in the inertial range. The dissipation range starts around $n \simeq 13 - 17$ for runs G1–G3; this is marked by the bend in the structure functions. The period-three oscillations mentioned earlier show clearly in the inertial range of figure 9. We have obtained ζ_p's by fitting the structure functions to a power-law form in the inertial range. As can be seen from figure 15, the fitting errors increase with increasing p because the period-three oscillations are stronger in high-order structure functions.

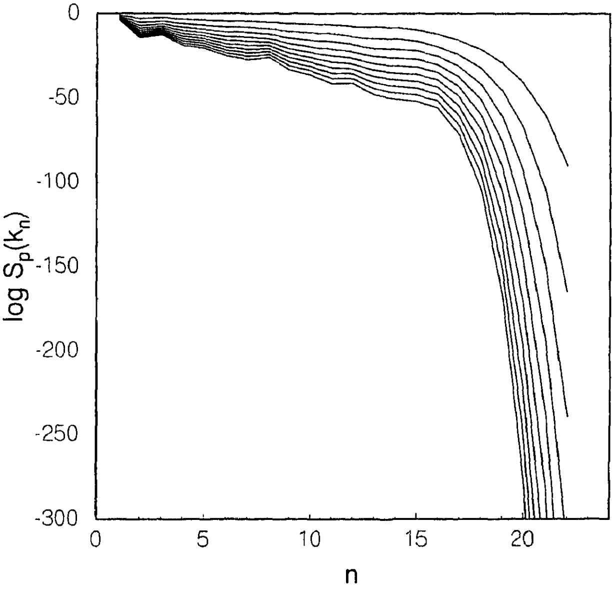


Figure 9. A plot of $\log S_p(k_n)$ versus the shell number n for the run G1 and for $p = 2, 4, 6, \ldots, 20$ (from top to bottom). Note the period-three oscillations that show clearly in high-order structure functions.

Kadanoff *et al* [84] have pointed out that a simple way of removing these oscillations is to calculate triple moments of velocities; the most effective in removing these oscillations is the structure function

$$\Sigma_p(k_n) = \left\langle \left| \Im\left(v_n v_{n+1} v_{n+2} + \frac{1-\delta}{\lambda} v_{n-1} v_n v_{n+1} \right) \right|^{p/3} \right\rangle, \tag{37}$$

which scales like $S_p(k_n)$ in the inertial range. Here $\Im$ denotes the imaginary part and we use $\delta = 1/2$ and $\lambda = 2$. In figure 10, we show plots of $\log \Sigma_p(k_n)$ against the shell number (n) for our run G1 with $2 \leq p \leq 20$; note that the period-three oscillations of $S_p(k_n)$ (figure 9) are absent and the inertial scaling region is a few octaves larger than the one in figure 9. A comparison of the exponents ζ_p calculated from $S_p(k_n)$ and $\Sigma_p(k_n)$ is shown in figure 11. The latter not only yields estimates with smaller error bars but also results much closer to the SL prediction (17). Thus we use $\Sigma_p(k_n)$ to calculate the Re_λ dependence of ζ_p (see below). However, for consistency we use $S_p(k_n)$ when we compare our GOY-model results with those of our direct numerical simulation of the 3*d* NS equation.

Before presenting our study of $\delta\zeta_p$ we note that a similar study has been carried out by using the hierarchical shell model studied by Eggers and Grossmann [95] and Grossmann and Lohse [96]. Their scheme is also referred to as the reduced wavevector approximation (REWA). The model considers the velocity field associated with the n-th shell to be given by a set of variables $v_{n,j}$, where $j = 1, \ldots, J^n$, with $J > 1$. Each variable $v_{n,j}$ interacts with all the variables in the $n \pm 1$ and $n \pm 2$ shells. The resulting set of

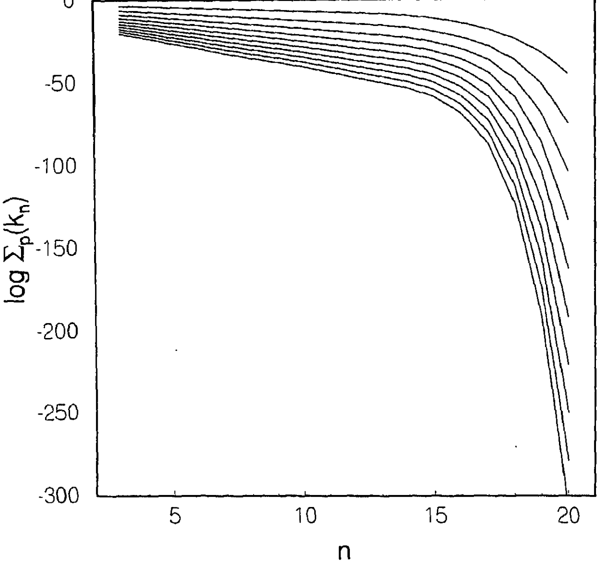


Figure 10. A plot of $\log \Sigma_p(k_n)$ (see text) versus the shell number n for the run G1 and for $p = 2, 4, 6, \ldots, 20$ (from top to bottom). Note that the curves are free from the oscillations that appear in the structure-function plots (figure 9).

differential equations has many more nonlinear terms than its analogue in the GOY model. The advantage of this added complexity is that the nonlinear term has a greater similarity with its counterpart in the NS equation (26). The REWA model has many more degrees of freedom than the GOY model, so its numerical solution is almost as difficult as a direct solution of the $3d$ NS equation. Of course, much higher values of Re can be achieved for the REWA models than for the $3d$ NS equation. We refer the reader to refs [8, 95, 96] for details of this model. Grossmann, Lohse, L'vov and Procaccia [35] have used the REWA model to cover the range $10^4 < \mathrm{Re} < 10^7$. (In these simulations, $\mathrm{Re}_\lambda \sim 12\sqrt{\mathrm{Re}}$ as shown in ref. [96].) From their plots of $\delta\zeta_p = |\zeta_p - p/3|$ versus Re, these authors argue that the theoretical prediction of ref. [36], namely, $\delta\zeta_p \sim \mathrm{Re}^{-3/10}$, might well be true as $\mathrm{Re} \to \infty$.

We have tried to address this issue for the GOY shell model. Specifically, we have obtained ζ_p and hence $\delta\zeta_p$ from $\Sigma_p(k_n)$ for eight different Reynolds numbers (runs G1-8, table 1) in the range $4 \times 10^4 \lesssim \mathrm{Re}_\lambda 3 \times 10^6$. The resulting data are shown in the plot of $\delta\zeta_p$ versus Re_λ in figure 12; the SL prediction is also shown for some values of p. There seems to be no indication of a decrease in $\delta\zeta_p$ with increasing Re_λ; if anything, there is a marginal increase. (We have checked that the increase in $\delta\zeta_p$ at the highest value of Re_λ (run G8) remains on choosing a smaller time step and larger averaging time for this run.) This result is in contrast with the conclusion of Grossmann *et al* [96]. It is possible that this discrepancy arises because of some feature that is present in the REWA model but not in the GOY shell model. Only more exhaustive numerical simulations (ideally of the $3d$ NS equation) and systematic experimental studies will be able to decide whether the trend shown in figure 12 actually obtains in fluid turbulence. We also note that, on the basis of their static analysis of the GOY model, Schörghofer, Kadanoff, and Lohse [92] have

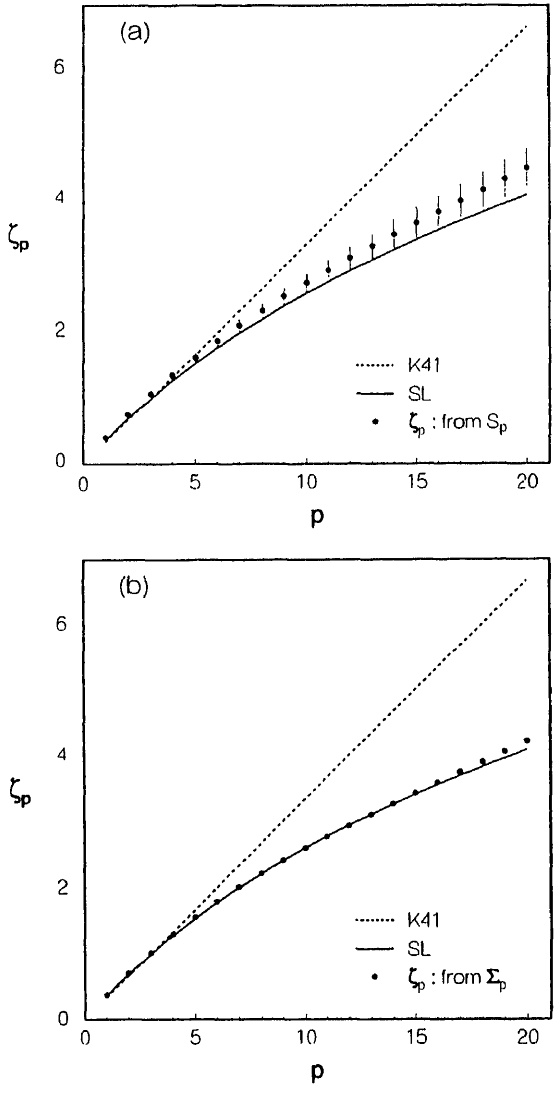


Figure 11. ζ_p versus p for our run G1 obtained from the best straight-line fit in the inertial range by using (**a**) $S_p(k_n)$ and (**b**) $\Sigma_p(k_n)$. Error bars in (**a**) denote the fitting errors; similar error bars are also shown in (**b**) but are not apparent since they are comparable to the symbol sizes. The K41 and the SL predictions are also shown.

suggested that ζ_p should have an oscillatory dependence on $\log \nu$; our fully dynamical calculation (figure 12) seems to be in agreement with this suggestion.

The obvious question that arises is whether the ζ_p's are universal. The data of figure 12 suggest they are not. This has also been found by Leveque and She [97] in their study of the GOY model in the presence of pure hyperviscosity, i.e., with a $-\nu_H k^{\alpha_H}$ dissipation and α_H not necessarily 2: their ζ_p's depend on α_H. Similar results have been obtained by

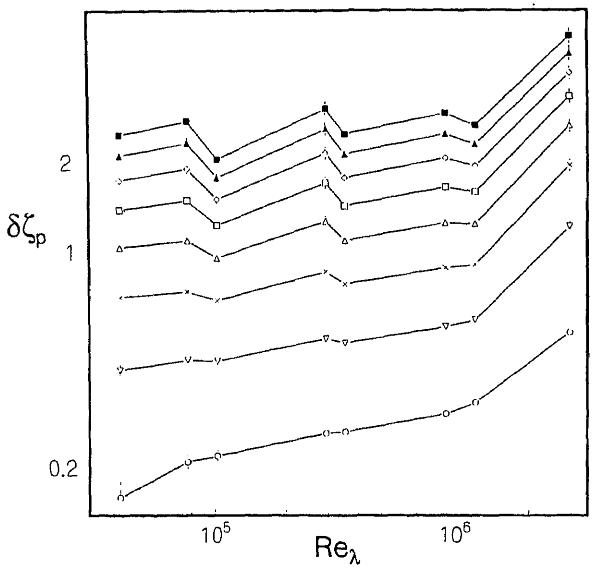


Figure 12. Log-log plot of $\delta\zeta_p$ versus the Taylor-microscale Reynolds number Re_λ for our GOY-model runs (G1-8) with $p = 6, 8, \ldots, 20$ (from bottom to top). The dotted ($p = 6$) and dashed ($p = 8$) lines show the SL results. Error bars are shown but are often smaller than the symbol sizes.

Borue and Orszag [98] in a direct numerical solution of the NS equation with pure hyperviscosity. However, this nonuniversality is removed [97] if one uses the ratios ζ_p/ζ_3. We shall return to this point in our discussion of k-space ESS and GESS below.

Before presenting our numerical studies of ESS and GESS, we wish to discuss the form of $E(k)$ in the dissipation range. It has been suggested by some authors [60, 99] that the energy spectrum in the far-dissipation range is of the form

$$E(k) \sim f(k/k_\mathrm{d}) \exp[-c(k/k_\mathrm{d})^n], \tag{38}$$

where c is a constant, f is a weak function of k/k_d, and $1 \le n \le 2$. In their direct numerical simulation of the NS equation, Chen *et al* [59] find that the form

$$E(k) \sim k^\alpha \exp(-ck/k_\mathrm{d}) \tag{39}$$

is consistent with their data with $\alpha \simeq 3.3$ and $c \simeq 7.1$. This pseudo-spectral simulation uses 256^3 Fourier modes and has achieved the best numerical resolution of the far dissipation range of the $3d$ NS equation to date. Our low-Re_λ studies (table 1) also yield a dissipation range in which $E(k)$ fits such a form. We have also checked explicitly that our data for S_3 in the GOY model can be fit to the form

$$S_3 \sim k^\delta \exp(-ck/k_\mathrm{d}) \tag{40}$$

in the dissipation range (run G1) with $\delta = 3.80$ and $c/k_\mathrm{d} = 4.13 \times 10^{-4}$. (We note that such fits are difficult in the GOY model and may not be unique: given the logarithmic sampling of k space , we have only five points in five octaves of k.)

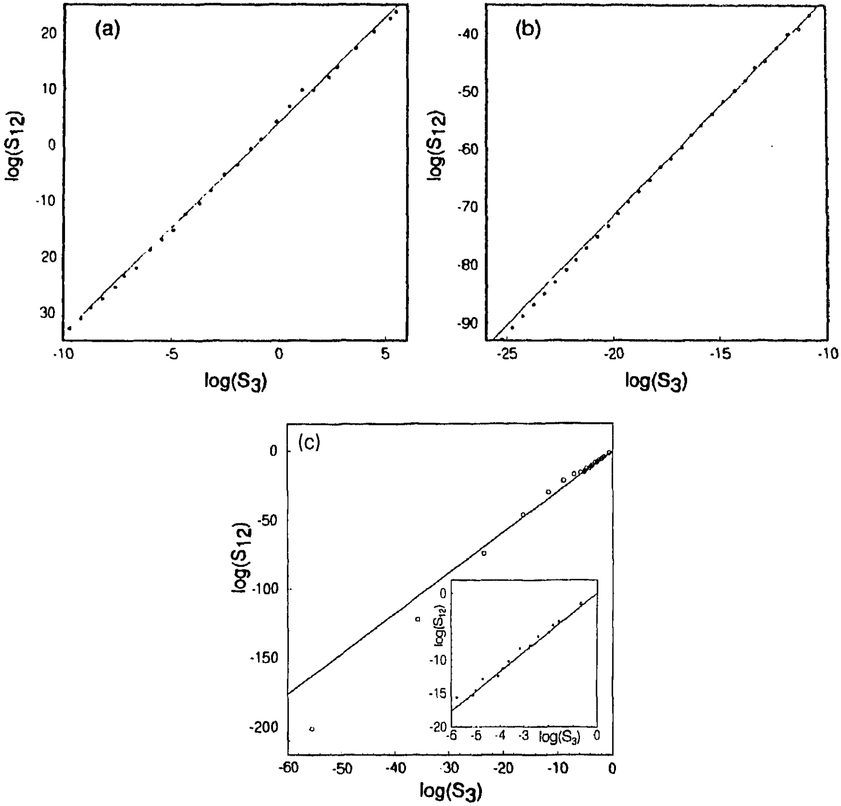


Figure 13. Log-log plots for $S_{12}(k)$ versus $S_3(k)$ for our runs (**a**) NS1 (the inertial range), (**b**) NS1 (the dissipation range) and (**c**) G1. The inset in (**c**) shows a magnified view of data in the inertial range. The straight lines represent $S_{12} \sim S_3^{\zeta_p^{\mathrm{SL}}}$. Deviations from this line show clearly for data in the dissipation range (large k, i.e., small S_3).

In recent work three of us [89] have proposed and verified a k-space version of ESS and GESS in the inertial range. We have also extended the applicability of ESS and GESS to the far dissipation range to uncover intriguing and apparently universal crossover from inertial to dissipation ranges. Our k-space ESS postulate is

$$
\begin{aligned}
S_p &\equiv \langle |\mathbf{v}(\mathbf{k})|^p \rangle \approx A_{Ip}(S_3)^{\zeta'_p}, \quad L^{-1} \ll k \lesssim 1.5k_{\mathrm{d}}, \\
S_p &\equiv \langle |\mathbf{v}(\mathbf{k})|^p \rangle \approx A_{Dp}(S_3)^{\alpha_p}, \quad 1.5k_{\mathrm{d}} \lesssim k \ll \Lambda,
\end{aligned}
\tag{41}
$$

where the A_{Ip} and A_{Dp} are, respectively, amplitudes for inertial and dissipation ranges and Λ^{-1} the (molecular) length at which hydrodynamics fails. Note that our postulate suggests different power-law behaviours in inertial and dissipation ranges; the real-space analogue of this was noted by Stolovitzky and Sreenivasan [22], who used experimental data, soon after the proposal of real-space ESS by Benzi, *et al* [32]. We do find that this way of presenting the data extends the apparent inertial range in both our NS and GOY-model studies, so the expression k–space ESS is justified. Representative log-log ESS

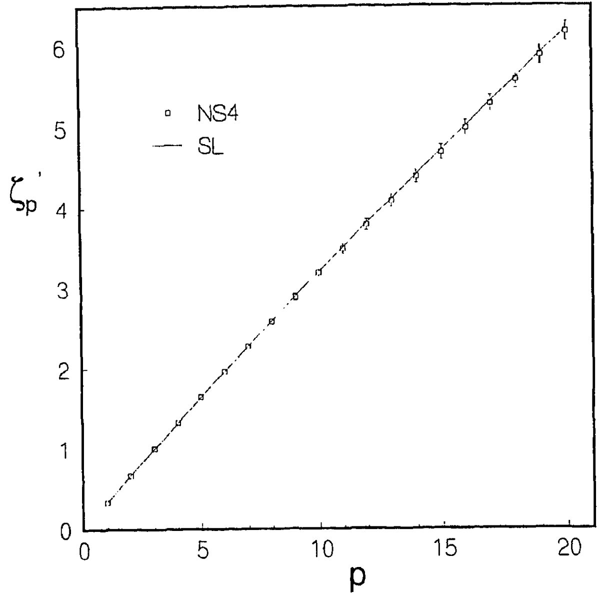


Figure 14. A plot of ζ_p' versus p. The squares indicate data from our run NS4. The line is $\zeta_p' = (2\zeta_p + 3p)/11$, with $\zeta_p = \zeta_p^{\mathrm{SL}}$ (see text).

plots of $S_{12}(k)$ versus $S_3(k)$ from our run NS1 (table 1) are shown in figure 13: here panel (a) covers the inertial range and yields the inertial-range power ζ_{12}', which is in agreement with the SL prediction; panel (b) covers the dissipation-range and yields a power α_{12}, which deviates noticeably from the SL prediction; panel (c) illustrates the same crossover from inertial- to dissipation-range asymptotic behaviours for the GOY shell model, where we can resolve much larger inertial and dissipation ranges, so this crossover shows clearly. Note also that the period-3 oscillations in figure 9 (structure functions for the GOY model) are removed in such ESS plots, yielding better estimates for the exponents. The agreement between our NS and GOY results, both qualitative and quantitative (see below), indicates that this crossover is real and not a numerical artifact. To ensure that we have good statistics, all our NS runs have been done in quadruple precision and we have checked that our estimates for ζ_p' and α_p do not change noticeably when we double our averaging time or halve our integration time step. Our NS runs are such that NS1, NS2, and NS3 resolve enough of the dissipation range (table 1) to yield dissipation-range exponents α_p; run NS4 does not yield many points in the dissipation range, so we do not use it to obtain α_p.

In (41) the inertial-range exponent: $\zeta_p' = \zeta_p$, for the GOY model; however, we find explicitly (figure 14) that, for the 3d NS case,

$$\zeta_p' = (\zeta_p + 3p/2)/(\zeta_3 + 9/2), \tag{42}$$

whence we get (36), i.e., $S_p(k) \sim k^{-(\zeta_p + 3p/2)}$, in the inertial range. To our knowledge this result for $S_p(k)$ is new. Our NS runs, though restricted to relatively low Re_λ ($\lesssim 22$), obtain it via ESS (41) and $\zeta_3 = 1$. For even values of p we can obtain this result via

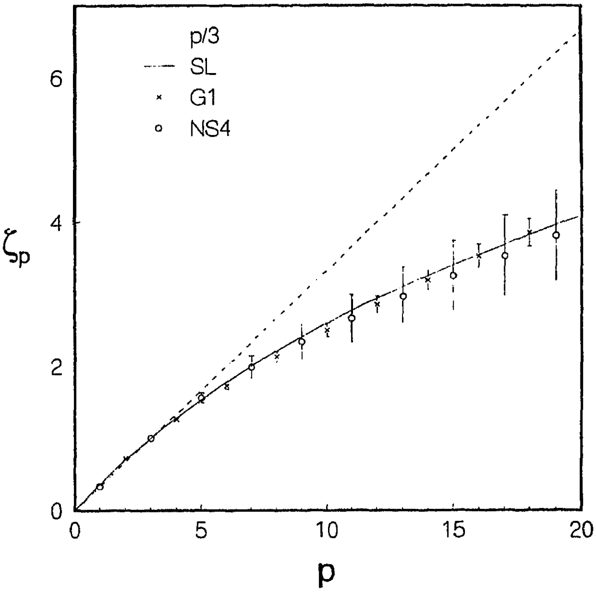


Figure 15. A plot of ζ_p versus p for the runs NS4 and G1. The K41 and SL lines are also shown. For NS4 we obtain ζ_p by inverting (47) and using our data for ζ'_p; this inversion also magnifies the error bars (cf. figure 14).The estimates for these exponents are obtained from ESS plots like figure 13 for both GOY and NS runs.

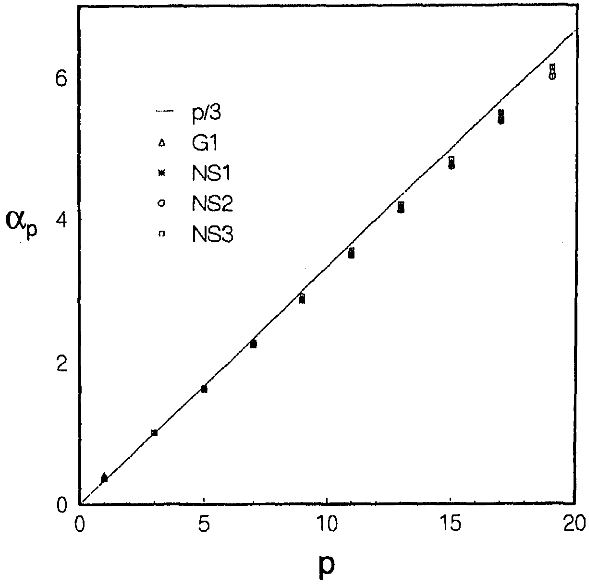


Figure 16. A plot of α_p versus p for the runs NS1-3 and G1. The K41 line is also shown. The estimates for α_p are obtained from the dissipation-range asymptotes of ESS plots like figure 13.

dimensional analysis if we Fourier transform the real-space structure function $\mathcal{S}_p$ and make one numerically plausible *assumption*: namely, that $\langle \mathbf{v}_{i1}(\mathbf{k}_1),\ldots,\mathbf{v}_{ip}(\mathbf{k}_p)\rangle$ is dominated by terms in which the $\mathbf{k}_m$, $m=1,\ldots,p$, arguments form equal and opposite pairs *all* with magnitude k, i.e., $\langle \mathbf{v}_{i1}(\mathbf{k}_1),\ldots,\mathbf{v}_{ip}(\mathbf{k}_p)\rangle \sim S_p(k)\ \ [\delta(\mathbf{k}_1+\mathbf{k}_2)\cdots\delta(\mathbf{k}_{p-1}+\mathbf{k}_p)+\textit{permutations}]$.

Both ζ_p (figure 15) and α_p (figure 16) seem universal, to the extent that, given our error bars, they are the same for all our GOY-model and $3d$ NS runs (table 1). ζ_p agrees reasonbly well with the She-Leveque (SL) formula [51] for the ranges of p and Re_λ in figure 15; and α_p is close to, but *systematically less* than, $p/3$. Of course, figure 12 indicates that the GOY-model exponents have a mild depedence on ν; however, as noted above, this nonuniversality is removed in our ESS and GESS plots (see below) in which only ratios of the exponents ζ_p appear.

Given our k-space ESS (41), the asymptotic k dependences of $S_p(k)$ in inertial and dissipation ranges follow from the dependence of S_3 on k. We find

$$S_3 \approx B_I k^{-\zeta_3-9/2}, \quad L^{-1} \ll k \lesssim 1.5k_{\mathrm{d}}, \tag{43}$$

$$S_3 \approx B_D k^{\delta}\exp(-ck/k_{\mathrm{d}}), \quad 1.5k_{\mathrm{d}} \lesssim k \ll \Lambda, \tag{44}$$

where B_I and B_D are, respectively, nonuniversal amplitudes. (Equation (44) holds for the $3d$ NS equation (see above); for the GOY model the factor $9/2$ is absent.) This implies that, in the far dissipation range,

$$S_p \sim k^{\theta_p}\exp(-c\alpha_p k/k_{\mathrm{d}}), \quad k_{\mathrm{d}} \ll k \ll \Lambda, \tag{45}$$

with $\theta_p = \alpha_p\delta$, for *all* p. This form is not easy to verify numerically for large p, given the rapid decay at large k. To the best of our knowledge it has been suggested so far [59] only for S_2.

In (44), δ, c, and k_{d} are not universal. Our k-space ESS shows that there is a universal part to the inertial- to dissipation-range crossover. We now show that this universal crossover can be extracted best by a k-space version of GESS that three of us have proposed recently [89]. This holds for both the GOY model and the NS equation. (Real-space GESS, due to Benzi *et al* [32], has been discussed in § 2.) We begin by defining

$$G_p \equiv S_p/(S_3)^{p/3}. \tag{46}$$

Log-log plots of G_p versus G_q yield curves with asymptotes which have *universal, but different,* slopes in inertial and dissipation ranges. Figure 17 shows a representative plot, with both GOY-model and NS data, for $p=6$ and $q=9$. The inertial-range asymptote has a slope $\rho(p,q)$ given by (21); it is easy to check from the formulae above that this is the same as the slope of real-space GESS plots [58] in the inertial range. From the SL value [51] for ζ_p (17) we can obtain the SL prediction for $\rho(p,q)$; the resulting inertial-range asymptote is in good agreement with our data in the inertial-range part of figure 17. From our k-space ESS and GESS ansätze above, it also follows that the dissipation-range asymptote has a slope

$$\omega(p,q) \equiv [\alpha_p - p/3]/[\alpha_q - q/3], \tag{47}$$

where α_p is defined in (41) and shown in figure 16. Note that GESS plots amplify

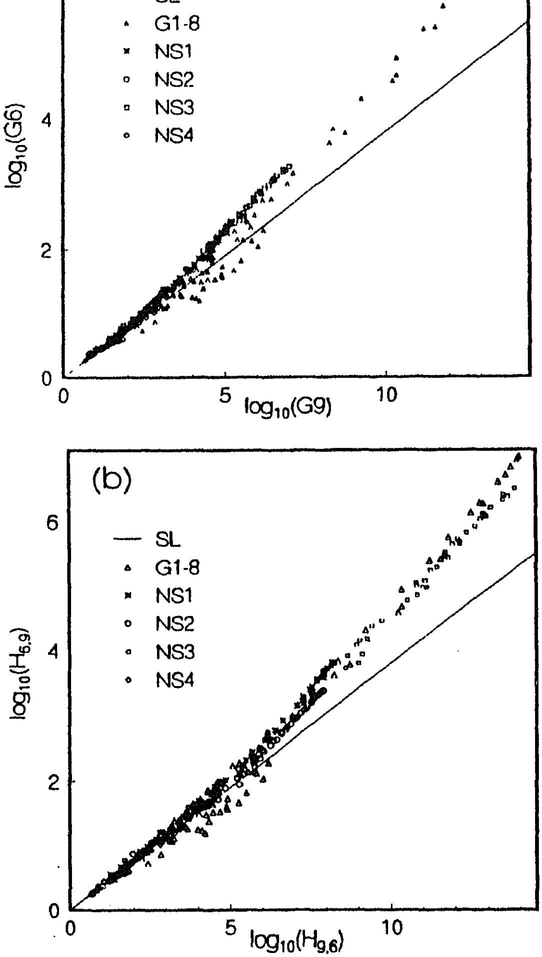


Figure 17. Log-log plots for (**a**) G_6 versus G_9, and (**b**) $H_{9,6}$ versus $H_{6,9}$ (see text) for all our NS and GOY-model runs. The data collapse in (**b**) yields the universal scaling function that characterizes the crossover from the inertial to the dissipation range, for $p = 6$ and $q = 9$ here.

considerably the difference between inertial- and dissipation-range asymptotes, which is not very prominent in ESS plots; this amplification can be estimated analytically.

In GESS plots like figure 17 the slopes of these inertial- and dissipation-range asymptotes are universal, at least at the level of accuracy of figures 15 and 16; but the point at which the curve in a GESS plot starts deviating from the inertial-range asymptote depends on the model (GOY, NS, with and without hyperviscosity, etc.). It is easily seen, though, that the following transformation yields a *universal crossover scaling function* (this curve is, of course, different for each pair of values of p and q because of multiscaling): We define

$$\log(H_{pq}) \equiv D_{pq} \log(G_p),$$
$$\log(H_{qp}) \equiv D_{qp} \log(G_q), \tag{48}$$

where the scale factors $D_{pq} = D_{qp}$ are *nonuniversal.* Now plots of $\log(H_{pq})$ versus $\log(H_{qp})$ show data collapsing onto *one universal curve* within our error bars. This is illustrated in figure 17b for $p = 6$ and $q = 9$; we emphasize that the data collapse occurs for *all* GOY and 3*d* NS runs and for *all* k and Re_λ. In essence this transformation holds one of the GESS plots fixed and slides those of other models till the asymptotes match.

In their discussion of real-space GESS, Benzi *et al* have suggested a hierarchy for the $\mathcal{G}_p(r)$, which leads to the SL formula if one uses the SL boundary conditions mentioned above. This hierarchy has a k-space analogue [89], namely

$$[G_{p+1}/G_p] = [G_p/G_{p-1}]^\gamma \times \left[\lim_{p\to\infty} G_{p+1}/G_p\right]^{1-\gamma}, \tag{49}$$

with $\gamma^3 = 2/3$. If we make the additional assumption $G_p(k) \sim k^{\beta_p}$, which holds in the inertial range, we get a difference equation for β_p *identical* to the SL one, which, when solved with the boundary conditions $\beta_0 = \beta_3 = 0$ and $\lim_{p\to\infty}(\beta_{p+1} - \beta_p) = 2/9$, yields the SL formula (17) via (7) (our $\beta_p = -\tau_{p/3}$ there). Elementary manipulations show that our k-space GESS gives $[G_{p+1}/G_p] \sim [G_p/G_{p-1}]^{\Upsilon_p}$, with $\Upsilon_p = (\zeta_{p+1} - \zeta_p - 1/3)/(\zeta_p - \zeta_{p-1} - 1/3)$. At first sight this seems inconsistent with the assumed hierarchy (49); however, it can be seen to be consistent if $\Upsilon_p = \gamma - 2(1-\gamma)/[9(\zeta_p - \zeta_{p-1} - \zeta_3/3)]$, which turns out to be the SL difference equation. Thus we cannot claim to have checked either of the assumed hierarchies (14) or (49) directly. Our k-space GESS can certainly hold with $\zeta_p \neq \zeta_p^{\mathrm{SL}}$. We can only claim that the hierarchy might hold to the extent that our calculated values of ζ_p agree reasonably with those predicted by the SL formula (figure 15).

The difference between inertial- and dissipation-range asymptotic behaviours has not been noted in real-space GESS so far. As we have mentioned above, experimental evidence for slope differences between inertial- and dissipation-range asymptotes in real-space ESS plots was given by Stolovitzky and Sreenivasan [22]. They postulated $\mathcal{S}_p \sim \mathcal{S}_3^{\alpha'_p}$ in the dissipation range and suggested

$$\alpha'_p \simeq \frac{\zeta_{3p/2} + p/2}{\zeta_{9/2} + 3/2}. \tag{50}$$

Unfortunately, we have not been able to find a simple, direct relation between their real-space exponents α'_p and our k-space exponents α_p (unlike (42) for inertial-range exponents), since S_p does not have a simple power-law dependence on k in the dissipation range. Furthermore, in this range, $G_p \sim k^{\beta'_p} \exp(-c'_p k/k_{\mathrm{d}})$, so there is no analogue of the SL formula for our dissipation-range exponents α_p. Nonetheless it should be interesting to study the apparently universal inertial- to dissipation-range crossover that we have elucidated above in experimental flows.

3.4 *Probability distributions*

Direct numerical simulations have obtained data for various probability ditributions [64, 73]. In these studies, which have attained $\mathrm{Re}_\lambda = 150$, it has been found that the distribution of velocity components is close to Gaussian. However, the distributions of velocity derivatives such as $\partial_x \mathrm{v}_x$ show significant deviations from Gaussian behaviour, which are consistent with experimental results. This deviation is even more visible in the distribution of derivatives like $\partial_y \mathrm{v}_x$. The PDFs of such velocity derivatives show tails that decay even more slowly than tails of exponential distributions.

PDFs of velocity differences have also been obtained by these direct numerical studies [73]. The distributions found are similar to those seen in experiments [31, 62] and seem to have exponential tails for small separtions r; as this separtion increases, the deviations from Gaussian character decrease till, at large separtions comparable to integral scales, these PDFs approach Gaussians.

These simulations [73] have tried to check the log-normal hypothesis of Kolmogorov [42] for the distribution of the energy dissipation rate ϵ. Though there is reasonable agreement near the peak of the distribution, there are significant deviations from this form in the tails. We are not aware of any direct numerical checks of the log-Poisson class of distributions that follow from the She–Leveque assumptions [55, 67]. Data from our Navier–Stokes runs would not be good enough to rule out other possible distributions. To obtain information about the distribution of ϵ in the GOY model, one must transform to real space. As we have said earlier, the logarithmic discretisation of k space makes such transformation questionable for this model, so we restrict ourselves to the distributions of Fourier components (see below). The best we can do is to check the analogue of She–Leveque hierarchy for the moments of ϵ, which implies a distribution of the log-Poisson class. We have discussed this above in terms of G_p (see the paragraph with (49)).

The study of probability distributions in shell models is of more recent origin. Non-Gaussian behaviour in the GOY shell model has been studied by Biferale [100]. He has shown, by taking a Fourier transform of v_n, that PDFs of the real-space velocity gradients are non-Gaussian for the GOY shell model and not far from the predictions of a multifractal theory. Also, the deviations from Gaussian behaviour increase with decreasing spatial separation, in qualitative accord with the results for the NS equation and from experiments. One might question whether Fourier transforms over the logarithmically spaced wavenumbers in the GOY model can give reasonable real-space information. Thus, in our study, we restrict ourselves to k-space quantities and look for possible non-Gaussian statistics. The REWA models sample k-space better than the GOY model. These have also been used to obtain the PDFs of velocity differences [101] which show a crossover from Gaussian behaviour, at the largest spatial scales, to PDFs with increasingly stretched tails at smaller scales.

In analogy with the experimental studies of Katsuyama *et al* we first construct k-space γ_{2p}'s (cf. (23)) for the GOY shell model:

$$\gamma_{2p} = \frac{(2p-1)!!}{2} \frac{\langle |v_n|^2 \rangle^p}{\langle |v_n|^{2p} \rangle}, \quad p = 2, 3, 4, \ldots, \tag{51}$$

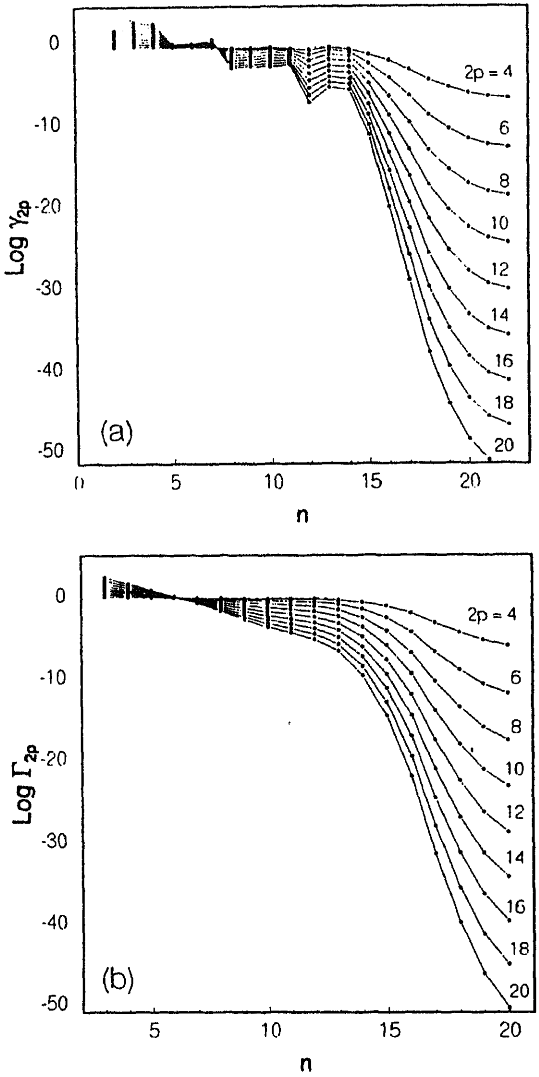


Figure 18. Plots of (**a**) γ_{2p} and (**b**) Γ_{2p} versus the shell number n for $2p = 4, 6, \ldots, 20$, for our run G1. The dashed lines indicate $\gamma_{2p} = 0.5$ and $\Gamma_{2p} = 0.5$, respectively. One can see that the plots in (**b**) are free from the oscillations present in (**a**).

which can be rewritten as

$$\gamma_{2p} = \frac{(2p-1)!!}{2} \frac{S_2^p}{S_{2p}} \quad p = 2, 3, 4, \ldots. \tag{52}$$

Note that the midband frequency f_c in (23) has been replaced, effectively, by the shell number n here. In figure 18a we plot γ_{2p} (calculated via eq. (52)) as a function of n, the shell number, for run G1. One can clearly observe the deviation of γ_{2p} from the Gaussian

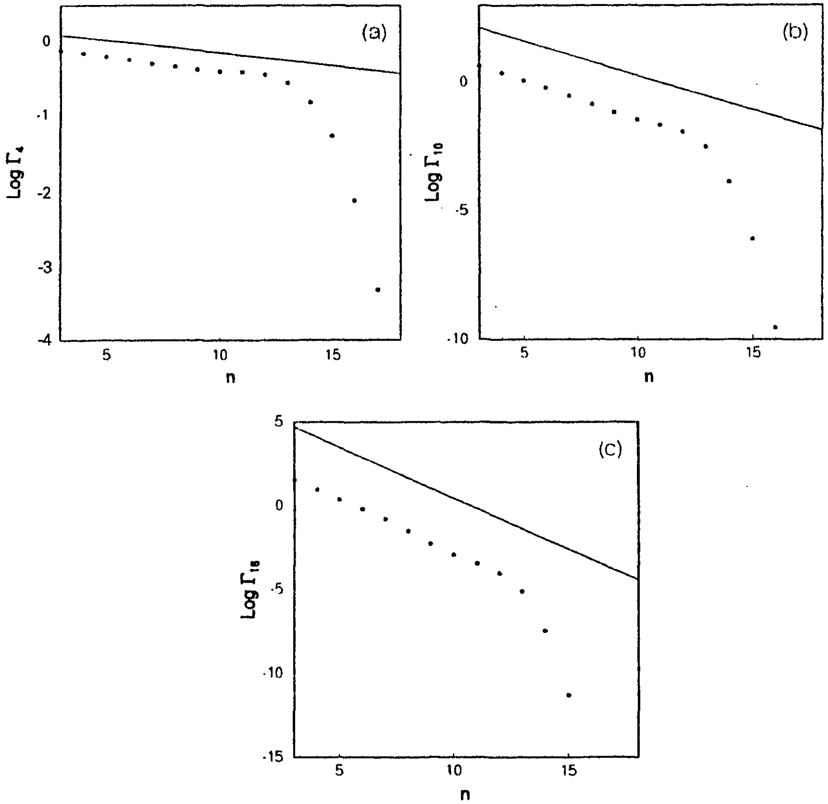


Figure 19. Inertial-range behaviour of (**a**) Γ_4 (**b**) Γ_{10} and (**c**) Γ_{16} for run G1 (figure 18). Continuous lines represent the form $\Gamma_{2p} \propto \frac{1}{2}(2p-1)!!\, k_n^{\zeta_{2p}-p\zeta_2}$, with the ζ_p's calculated from the SL formula.

value of 0.5 in the inertial range and the rapid decay of γ_{2p} in the dissipation range (beyond $n = 14$). However, the curves show oscillations that arises because of the underlying 3-cycle in the GOY shell model (see above). To eliminate these we define a new function Γ_{2p} in terms of Σ_p ((37)):

$$\Gamma_{2p} = \frac{(2p-1)!!}{2} \frac{\Sigma_2^p}{\Sigma_{2p}}, \quad p = 2, 3, 4, \ldots. \tag{53}$$

Figure 18b shows the variation of Γ_{2p} (calculated using (53)) with n for run G1. As one might have anticipated, the use of the correlation function Σ_p has removed the oscillations so the deviations from Gaussian behaviour show clearly.

Since we know that $\Sigma_p \sim k_n^{-\zeta_p}$ in the inertial range, we can infer from (53) the form of Γ_{2p} in the inertial range, i.e.,

$$\Gamma_{2p} \propto \frac{(2p-1)!!}{2} k_n^{\zeta_{2p}-p\zeta_2}, \quad L^{-1} \ll k \lesssim 1.5 k_{\rm d}. \tag{54}$$

To specify Γ_{2p} completely in the inertial range we must also provide the amplitudes for the power-law dependence of Σ_p on k. However, this is not necessary for the purpose

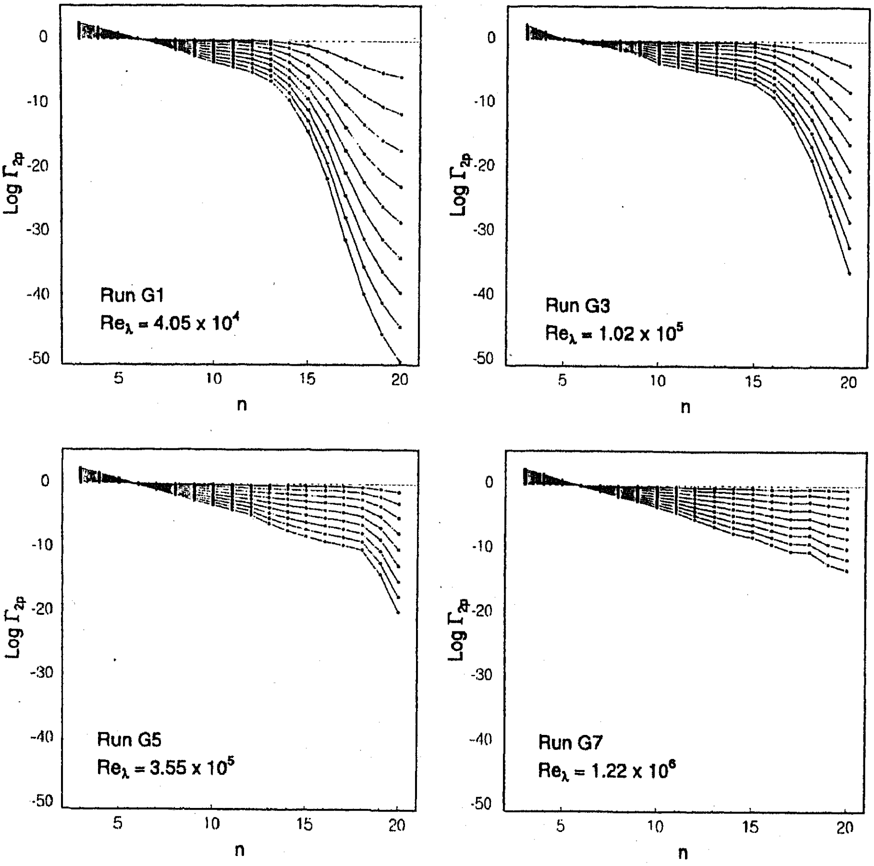


Figure 20. Γ_{2p} as functions of the shell number n for our runs G1, G3, G5, and G7 for $2p = 4, 6, \ldots, 20$ (top to bottom). Dashed lines indicate $\Gamma_{2p} = 0.5$.

of figure 19, where we show comparisons between the Γ_{2p}'s shown in figure 18b and the form suggested in (54) (with the ζ_p's calculated from the SL formula) for $2p = 4, 10$ and 16. Clearly, similar power laws obtain in the inertial range. We note in passing that the slight hump before the dissipation range is an indication of the bottleneck phenomenon [61].

Figure 20 shows Γ_{2p} as a function of the shell number n for four different Re_λ (our runs G1, G3, G5 and G7). The general conclusion that we can draw from these graphs is that the distribution $P(|v_n|)$ crosses over from being close to Gaussian at small k_n to ones that fall more and more slowly with increasing k_n. This is in accord with the data of Katsuyama *et al* (1994) (note our wavenumber k_n is the analogue of their midband frequency f_c). These graphs also illustrate that, with increasing Re_λ, the extent of the inertial range increases and that the deviations from Gaussian distributions is far more in the dissipation range than in the inertial range. Direct plots of $\log \Gamma_{2p}$ versus Re_λ are in general agreement with these statements, but are not very smooth (like our plots of $\delta\zeta_p$ versus Re_λ (figure 12)) since we have studied only eight widely separated values of Re_λ.

4. Concluding remarks

We have provided a brief overview of experimental and numerical studies of coherent structures, structure functions, and probability distributions in turbulent flows that are, to a good approximation, homogeneous and isotropic. We have also outlined some of the phenomenological and stochastic models that have been used to develop an understanding of these quantities. However, as is well known, a complete theoretical understanding of homogeneous, isotropic turbulence has remained an elusive goal, in spite of the substantial progress that has been made in all the areas described above.

Our emphasis here has been on using direct numerical studies of the deterministically forced NS equation and the GOY shell model in conjunction with each other. In particular, we have explored, via ESS and GESS, an apparently universal crossover from inertial- to dissipation-range behaviours in these models. Thus our study adds to the growing evidence for similar multiscaling in NS and GOY models. It is tempting to say, therefore, that the GOY model captures some of the universal scaling properties of fluid turbulence because all known conservation laws have been built in by a suitable choice of parameters (see above). However, we must temper this optimistic view for there are many features, which seem important for fluid turbulence, that are not contained in, or poorly represented by, the GOY model. As noted by Kadanoff *et al* [84], there is no analogue of sweeping effects in the GOY model. Furthermore, since the GOY model has only scalar velocities and k's which are logarithmically spaced, it cannot represent well the high-vorticity filamentary structures, which are believed to be important in fully developed turbulence. Recall that the arguments which lead to the She–Leveque formula rely on the filamentary nature of these structures; yet, strangely enough, the GOY-model exponents ζ_p agree reasonably well with this formula. Are all these apparently important effects, which are not contained in the GOY shell model, *irrelevant* in some way? Unfortunately there is no clear answer to this question at this moment. Further studies are needed to elucidate the similarities and differences between the statistical properties of the solutions of the GOY shell model and the $3d$ NS equation. However, given that numerical studies of the GOY shell model are clearly far easier than direct numerical simulations of the $3d$ NS equation, the former can be used (with caution) as preliminary testing grounds for new theoretical ideas.

Acknowledgements

The authors thank C Jayaprakash, R Narasimha and S Ramaswamy for discussions and the referee for helpful suggestions. They also thank CSIR and BRNS, India for support and SERC, Bangalore for computational resources.

References

[1] M Van Dyke, *An album of fluid motion* (The Parabolic Press, Stanford, California, 1982)
[2] K R Sreenivasan and R A Antonia, *Ann. Rev. Fluid Mech.* **29**, 435 (1997)
[3] E D Siggia, *Ann. Rev. Fluid Mech.* **26**, 137 (1994)
[4] M Nelkin, *Adv. Phys.* **43**, 143 (1994)
[5] V L'vov and I Procaccia, *Phys. World* **35**, (1996)

[6] W D McComb, *The physics of fluid turbulence*, (Oxford University Press, Oxford, 1991)
[7] U Frisch, *Turbulence: the legacy of A N Kolmogorov* (Cambridge University Press, Cambridge, 1995)
[8] T Bohr, M H Jensen, G Paladin, A Vulpiani, *Dynamical systems approach to turbulence*, to be published (Cambridge University Press, Cambridge, 1997)
[9] G K Batchelor, *The theory of homogeneous turbulence* (Cambridge University Press, Cambridge, 1953)
[10] H Tennekes and J L Lumley, *A first course in turbulence* (MIT Press, Cambridge, Massachusetts, 1972)
[11] A S Monin and A M Yaglom, *Statistical fluid mechanics* (MIT Press, Cambridge, Massachusetts, 1975)
[12] G I Taylor, *Proc. R. Soc. (London)* **A151**, 421 (1935)
[13] G S Saddoughi and S V Veeravalli, *J. Fluid Mech.* **268**, 333 (1994)
[14] N Goldenfeld, *Lectures on phase transitions and the renormalization group* (Addison-Wesley, New York, 1992)
[15] To the best of our knowledge, a direct mapping of a deterministic partial differential equation (PDE) with spatiotemporal chaos onto a stochastic PDE has been carried out only for the Kuramoto–Sivashinsky (KS) equation. This mapping uses a numerical coarse-graining procedure and shows, both in one and two spatial dimensions, that the KS equation is in the universality class of the Kardar–Parisi–Zhang (KPZ) equation, i.e., the long-distance and long-time behaviours of their correlations functions are the same. (See: S Zaleski, *Physica* **D34**, 427 (1989); F Hayot, C Jayaprakash and Ch Josserand, *Phys. Rev.* **E47**, 911 (1993); C Jayaprakash, F Hayot and R Pandit, *Phys. Rev. Lett.* **71**, 15 (1993).) The case of the deterministically forced Navier–Stokes equation is considerably more subtle; however, in the absence of a direct mapping, it has been conjectured that the appropriate PDE is the NS equation with an additive, Gaussian white noise whose variance has a power-law dependence on the wave-vector [16,17,18]
[16] C DeDominicis and P C Martin, *Phys. Rev.* **A19**, 419 (1979)
[17] V Yakhot and S A Orszag, *J. Sci. Comput.* **1**, 3 (1986); *Phys. Rev. Lett.* **57**, 1722 (1986)
[18] J K Bhattacharjee, *J. Phys.* **A21**, L 551 (1988); *Phys. Rev.* **A40**, 6374 (1989); *Phys. Fluids* **A3**, 879 (1991)
[19] C-Y Mou and P B Weichman, *Phys. Rev. Lett.* **70**, 1101 (1993)
[20] J K Bhattacharjee, *Pramana – J. Phys.* **48**, 365 (1997)
[21] P Bak, C Tang and K Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987); *Phys. Rev.* **A38**, 364 (1988)
[22] G Stolovitzky and K R Sreenivasan, *Phys. Rev.* **E48**, R33 (1993)
[23] S Douady, Y Couder and M E Brachet, *Phys. Rev. Lett.* **67**, 983 (1991)
[24] O Cadot, S Douady and Y Couder, *Phys. Fluids* **7**, 630 (1995)
[25] E Villermaux, E Sixou and Y Gagne, *Phys. Fluids* **7**, 2008, (1995)
[26] L F Richardson, *Weather prediction by numerical process* (Cambridge University Press, Cambridge, 1922)
[27] A N Kolmogorov, *C.R. Acad. Sci. USSR* **30**, 301 (1941)
[28] N Cao, S Chen and K R Sreenivasan, *Phys. Rev. Lett.* **77**, 3799 (1996)
[29] H L Grant, R W Stewart and A Moilliet, *J. Fluid Mech.* **12**, 241 (1962)
[30] K R Sreenivasan, *Phys. Fluids* **7**, 2778 (1995)
[31] F Anselmet, Y Gagne, E J Hopfinger and R A Antonia, *J. Fluid Mech.* **140**, 63 (1984)
[32] R Benzi, S Ciliberto, R Trippiccione, C Baudet, F Massaioli and S Succi, *Phys. Rev.* **E48**, R29 (1993)
[33] J Herweijer and W van de Water, *Phys. Rev. Lett.* **74**, 4651 (1995)
[34] T Katsuyama, Y Horiuchi and K Nagata, *Phys. Rev.* **E49**, 4052 (1994)
[35] S Grossman, D Lohse, V L'vov and I Procaccia, *Phys. Rev. Lett.* **73**, 432 (1994)
[36] V S L'vov and I Procaccia, *Phys. Rev. Lett.* **74**, 2690 (1994)
[37] G I Barenblatt and N Goldenfeld, *Phys. Fluids* **7**, 3078 (1995)
[38] G Zocchi, J Maurer, P Tabeling and H Williame, *Phys. Rev.* **E50**, 3693 (1994)
[39] B Chabaud, A Naert, J Peinke, F Chilla, B Castaing and B Hebral, *Phys. Rev. Lett.* **73**, 3227 (1994)

[40] V Emsellem, L P Kadanoff, D Lohse, P Tabeling and J Wang, chao-dyn/9604009, *Phys. Rev. E* to appear

[41] A N Kolmogorov, *J. Fluid Mech.* **13**, 82 (1962)

[42] A A Praskovsky, *Phys. Fluids* **A4**, 2589 (1992)

[43] G Stolovitzky, P Kailasnath and K R Sreenivasan, *J. Fluid Mech.* **297**, 275 (1995)
G Stolovitzky and K R Sreenivasan, *Rev. Mod. Phys.* **66**, 229 (1994)

[44] S T Thoroddsen and C W Van Atta, *Phys. Fluids* **A4**, 2592 (1992)

[45] Y Gagne, M Marchand and B Castaing, *J. Phys.* **4**, 1 (1994)

[46] S Chen, G D Doolen, R H Kraichnan and L P Wang, *Phys. Rev. Lett.* **74**, 1755 (1995)

[47] K R Sreenivasan and P Kailasnath, *Phys. Fluids* **A5**, 512 (1993)

[48] U Frisch, P L Sulem and M Nelkin, *J. Fluid Mech.* **87**, 719 (1978)

[49] G Parisi and U Frisch, in *Turbulence and predictability in geophysical fluid dynamics* edited by M Ghil, R Benzi and G Parisi (North-Holland, Amsterdam, 1985) pp. 84–87

[50] C Meneveau and K R Sreenivasan, *J. Fluid Mech.* **224**, 429 (1991)

[51] Z S She and E Leveque, *Phys. Rev. Lett.* **72**, 336 (1994)

[52] E D Siggia, *J. Fluid Mech.* **107**, 375 (1981)

[53] Z S She, E Jackson and S A Orszag, *Nature* (*London*) **344**, 226 (1990)

[54] G R Chavarria, C Baudet and S Ciliberto, *Phys. Rev. Lett.* **74**, 1986 (1995)

[55] B Dubrulle, *Phys. Rev. Lett.* **73**, 959 (1994)

[56] D Segel, V L'vov and I Procaccia, *Phys. Rev. Lett.* **76**, 1828 (1996)

[57] C Meneveau, *Phys. Rev.* **E54**, 3657 (1996)

[58] R Benzi, L Biferale, S Ciliberto, M Struglia and R Tripiccione, *Europhys. Lett.* **32**, 709 (1995)

[59] S Chen, G Doolen, J R Herring, R H Kraichnan, S A Orszag and Z S She, *Phys. Rev. Lett.* **70**, 3051 (1993)
S Chen, G D Doolen, R H Kraichnan and Z S She, *Phys. Fluids* **A5**, 458 (1993)

[60] R H Kraichnan, *J. Fluid Mech.* **5**, 497 (1959)

[61] D Lohse and A Müller-Groeling, *Phys. Rev. Lett.* **74**, 1747 (1995); *Phys. Rev.* **E54**, (1996)

[62] A Praskovsky and S Oncley, *Phys. Rev. Lett.* **73**, 3399 (1994)

[63] R Benzi, L Biferale, G Paladin, A Vulpiani and M Vergassola, *Phys. Rev. Lett.* **67**, 2299 (1991)

[64] Z S She, *Phys. Rev. Lett.* **66**, 600 (1991)

[65] P Kailasnath, K R Sreenivasan and G Stolovitzky, *Phys. Rev. Lett.* **68**, 2766 (1992)

[66] G Stolovitzky, *The statistical order of small scale turbulence*, Ph.D. Thesis, Yale University, USA (1994), cited in ref. [2]

[67] Z S She and E C Waymire, *Phys. Rev. Lett.* **74**, 262 (1995)

[68] S A Orszag and G S Patterson, *Phys. Rev. Lett.* **28**, 76 (1972)

[69] R M Kerr, *J. Fluid Mech.* **153**, 31 (1985)

[70] M M Rogers and P Moin, *J. Fluid Mech.* **176**, 33 (1987)

[71] K Yamamoto and I Hosokawa, *J. Phys. Soc. Jpn.* **57**, 1532 (1988)

[72] R H Kraichnan and R Panda, *Phys. Fluids* **31**, (1988)

[73] M Meneguzzi and A Vincent, in *Advances in turbulence 3* edited by A V Johansson and P H Alfredsson (Springer, Berlin, 1991) pp. 211–220; A Vincent and M Meneguzzi, *J. Fluid Mech.* **258**, (1994)

[74] S Kida and K Ohkitani, *Phys. Fluids* **A4**, 1018 (1992)

[75] J Jimenez, A A Wray, P G Saffman and R S Rogallo, *J. Fluid Mech.* **255**, 65 (1993)

[76] N Cao, S Chen, and Z-S She, *Phys. Rev. Lett.* **77**, 3711 (1996)

[77] E B Gledzer, *Sov. Phys. Dokl.* **18**, 216 (1973)

[78] K Ohkitani and M Yamada, *Prog. Theor. Phys.* **81**, 329 (1989)

[79] A M Obukhov, *Atmos. Oceanic Phys.* **7**, 41 (1971)

[80] V N Desnyansky and E A Novikov, *Atmos. Oceanic Phys.* **10**, 127 (1974)

[81] M Yamada and K Ohkitani, *J. Phys. Soc. Jpn.* **56**, 4210 (1987)

[82] M H Jensen, G Paladin and A Vulpiani, *Phys. Rev.* **A43**, 798 (1991)

[83] D Pisarenko, L Bieferale, D Courvoisier, U Frisch and M Vergassola, *Phys. Fluids* **A5**, 2533 (1993)

[84] L Kadanoff, D Lohse, J Wang and R Benzi, *Phys. Fluids* **7**, 617 (1995)
[85] Sujan K Dhar, Ph.D. thesis, Indian Institute of Science, Bangalore (1996) unpublished
[86] L Kadanoff, D Lohse and N Schörghofer, *Physica* **D160**, 165 (1997)
[87] L Biferale, A Lambert, R Lima and G Paladin, *Physica* **D80**, 105 (1995)
[88] Ohkitani and Yamada [79] have shown that the quadratic invariant $\sum(-k_n)^\alpha|v_n|^2$ of order α is conserved in the inviscid, unforced GOY shell model provided one chooses $\delta = 1 - \lambda^{-\alpha}$. If the invariant of order α is conserved, then all lower-order invariants are also conserved. The choice $\alpha = 1$, i.e., $\delta = 1/2$ for $\lambda = 2$ thus conserves both the GOY-model energy and helicity in the inviscid, unforced limit; $\alpha = 2$, i.e., $\delta = 5/4$ also conserves the GOY-model enstrophy in this limit also, which is of relevance in models for two-dimensional turbulence
[89] S K Dhar, A Sain and R Pandit, *Phys. Rev. Lett.* (to appear) (1997)
[90] L Biferale and R Kerr, *Phys. Rev.* **E52**, 61133 (1995)
[91] R Benzi, L Biferale and G Parisi, *Physica* **D65**, 163 (1993)
[92] N Schörghofer, L Kadanoff and D Lohse, *Physica* **D88**, 40 (1995)
[93] Z S She, E Jackson and S A Orszag, *Proc. R. Soc. London* **A434**, 101 (1991)
[94] M E Brachet, D I Meiron, S A Orszag, B G Nickel, R H Morf and U Frsich, *J. Fluid Mech.* **130**, 411 (1983)
[95] J Eggers and S Grossmann, *Phys Fluids* **A3**, 1958 (1991)
[96] S Grossmann and D Lohse, *Phys. Rev.* **E50**, 2784 (1994)
[97] E Leveque and Z S She, *Phys. Rev. Lett.* **75**, 2690 (1995)
[98] V Borue and S A Orszag, *Europhys. Lett.* **29**, 6875 (1995)
[99] C Foias, O Manley and L Sirovich, *Phys. Fluids* **A2**, 464 (1990)
[100] L Biferale, *Phys. Fluids* **A5**, 428 (1993)
[101] J Eggers and S Grossmann, *Phys. Lett.* **A156**, 444 (1991)
[102] D Lohse, *Phys. Rev. Lett.* **73**, 3223 (1994)

physics
pp. 365–378

The randomly stirred fluid – Turbulence as a problem in statistical mechanics

J K BHATTACHARJEE
Department of Theoretical Physics, Indian Association for Cultivation of Science, Calcutta 700 032, India

Abstract. Properties of the randomly stirred fluid and the relevance to the problem of homogeneous isotropic turbulence are discussed.

Keywords. Homogeneous isotropic turbulence; randomly stirred fluids.

PACS Nos 47.27; 05.70

1. Preliminaries

This article deals with applications of the methods of stastistical mechanics to the problem of homogeneous, isotropic turbulence. Turbulence [1–7] deals with a wide class of flows which seem to possess complex and seemingly random structure at some macroscopic scale of dynamic importance. The chief physical characteristic is a tremendous enhancement in transport properties—transport of momentum, energy and particles is orders of magnitude higher than possible by molecular process. Equally important is the sensitive dependence on initial conditions [8] of a turbulent flow. Two turbulent flows which are nearly identical at a given time do not remain so on timescales of dynamical interest. It is the sensitivity to initial conditions which makes the methods of statistical mechanics applicable to the problem of turbulence. The key to this is the fact that while details of fully developed turbulence are sensitive to triggering disturbances, average properties are not. Hence statistically averaged quantities are the dynamical variables of interest.

The governing dynamical equation for the study of turbulence is the Navier–Stokes equation for the velocity field $\mathbf{v}(\mathbf{r},\mathbf{t})$,

$$\frac{\partial \mathbf{v}}{\partial t}+(\mathbf{v}\cdot\nabla)\mathbf{v}=-\frac{\nabla P}{\rho}+\nu\nabla^2\mathbf{v} \tag{1.1}$$

where P is the pressure, ρ the density and ν the kinetic viscosity. For an incompressible fluid, there is the additional condition that

$$\nabla\cdot\mathbf{v}=0. \tag{1.2}$$

All the information regarding the turbulence should be contained in (1.1), when it is supplemented with the boundary conditions and initial conditions. We now demonstrate that maintained turbulent motion in a closed system cannot be described by (1.1). To do boundaries, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\int v^2\mathrm{d}^3r = -2\nu\int \mathrm{d}^3r\left[\frac{\partial v_\alpha}{\partial x_\beta}\frac{\partial v_\alpha}{\partial x_\beta}\right] \tag{1.3}$$

and since the rate of change is always negative, motion has to cease at $t = \infty$. Thus, it is necessary to augment the right hand side of (1.1), by a force f, which feeds energy into the system. We will endow this force f, in what follows, with statistical property.

Even if the flow is not bounded, it is natural to think of an additional force, if we are worrying about the dynamics of the fluctuating part of the velocity field. To see this, we decompose the total velocity field as

$$\mathbf{v}(\mathbf{r},t) = \mathbf{V}(\mathbf{r}) + \mathbf{u}(\mathbf{r},t) \tag{1.4}$$

where $\mathbf{V}$ is a mean flow and $\mathbf{u}$ is a fluctuating field of zero mean. In other words, $\langle\mathbf{v}(\mathbf{r},t)\rangle = \mathbf{V}(\mathbf{r})$, where the prescription for taking averages will have to be provided. If we insert the above decomposition into the Navier–Stokes equation for incompressible flow, then we arrive at

$$\frac{\partial u_i}{\partial t} = \nu\nabla^2 u_i + \nu\nabla^2 V - \frac{\partial\bar{P}}{\partial x_i} - \frac{\partial p}{\partial x_i} - \frac{\partial}{\partial x_j}(V_iV_j + u_iu_j + V_iu_j + V_ju_i) \tag{1.5}$$

(note that pressure has been decomposed as $\bar{P} + p$).

Statistical average yields

$$\nu\nabla^2 V - \frac{\partial\bar{P}}{\partial x_i} = \frac{\partial}{\partial x_j}(V_iV_j + \langle u_iu_j\rangle) \tag{1.6}$$

and thus

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_j}(u_iu_j) = -\frac{\partial p}{\partial x_i} + \nu\nabla^2 u_i + f_i, \tag{1.7}$$

where

$$f_i = -\frac{\partial}{\partial x_j}(V_iu_j + V_ju_i + u_iu_j - \langle u_iu_j\rangle). \tag{1.8}$$

The fluctuating velocity field thus experiences an extra force coming mainly from the mean flow. It is this interaction of the fluctuating part with the mean flow that sustains the turbulence.

Now, about the averaging process. The available averages are space averages, time averages and ensemble averages. Spatial averages make sense only if scales over which the turbulence is homogeneous or approximately so are much larger than the scale of turbulent fluctuations. Similarly, the time averaging is useful if the time over which the turbulence is statistically stationary is longer than the time for turbulent fluctuations. Experiments usually deal with time averages obtained by using a local probe. The third average is the ensemble average where the average is taken over different realizations of the turbulent field. This ensemble averaging makes sense and can be equal to the time turbulent flow strongly mixing and ergodic. However, this would imply the existence of a stable asymptotic stationary state. It is likely that such a state exists but its existence has not been proven.

We now return to (1.7) and try to specify the properties of **f**. To ignore the questions of boundary conditions, mean flow, initial conditions etc., we make **f** a random force. We then need to specify the correlations of **f**. To do this, we need to keep in mind that the role of f is to supply energy at large length scales—preferably near the boundaries. If $\bar{\epsilon}$ is the rate per unit time, per unit mass at which energy is injected into the system and k_0 is a small wavenumber ($k_0 \sim O(L^{-1})$), L being the system size), then a possible prescription for the correlation is (note, dimension of $\bar{\epsilon}$ is L^2/T^3, L and T being scales for length and time)

$$\langle f_i(\mathbf{k},\omega) f_j(\mathbf{k}',\omega')\rangle = \bar{\epsilon}\frac{P_{ij}(k)}{k^{D-2}}\delta(k^2-k_0^2)\delta^D(\mathbf{k}+\mathbf{k}')\delta(\omega+\omega'). \tag{1.9}$$

In the above $P_{ij}(k)$ is the projection operator which ensures that $\nabla\cdot\mathbf{f}=0\,[P_{ij}(k)=\delta_{ij}-(k_ik_j/k^2)]$, D is the dimensionality of space and the factor k^{D-2} is required to maintain the correct dimensions. We now use the approximation that instead of being strictly zero for $k>k_0$ as required by the δ-function in (1.9), the correlation will fall off as a power law $f(k)$ expressed by

$$f(k) = k^{2-y} \tag{1.10}$$

to write the correlation of (1.9) as

$$\langle f_i(\mathbf{k},\omega) f_j(\mathbf{k}',\omega')\rangle = \frac{D_0}{k^{D-4+y}}P_{ij}\delta^D(\mathbf{k}+\mathbf{k}')\delta(\omega+\omega') \tag{1.11}$$

with y taken to be an arbitrary parameter. We thus arrive at the randomly stirred model [9] of De Domincis and Martin [10, 11]. The question arises if this model is to describe fully developed turbulence, then is there a special value of y.

To address the issue of what value of y is relevant, we need to discuss correlation functions, energy spectrum, inertial range [12] etc. To begin with, let us consider the velocity correlation function $C_{ij}(k,\omega)$ defined as

$$\begin{aligned} C_{ij}(k,\omega) &= \frac{\langle v_i(\mathbf{k},\omega)v_j(\mathbf{k}',\omega')\rangle}{\delta^D(\mathbf{k}+\mathbf{k}')\delta(\omega+\omega')} \\ &= C(k,\omega)P_{ij}(k) \end{aligned} \tag{1.12}$$

The expectation value is taken over the probability distribution associated with the noise in the randomly stirred model introduced above. The total energy in the turbulence field is given by $\int \mathrm{d}^Dk\,\mathrm{d}\omega C(k,\omega)$ and the energy spectrum $E(k)$ is defined as

$$\int E(k)\mathrm{d}k = \int \mathrm{d}^Dk\,\mathrm{d}\omega\, C(k,\omega). \tag{1.13}$$

A dimensional analysis yields the scaling behaviour of $E(k)$ in what is known as the inertial range. The inertial range encompasses k-values which are neither too small, nor too big. The small k-values correspond to the length scales at which the forcing occurs and energy is injected into the system. The large k-values correspond to small length scales where the molecular viscosity is important and dissipates the injected energy. The result is a maintained stationary state where the energy flows from the large length scales to the short length scales at a constant rate, which is the $\bar{\epsilon}$ introduced before. The main assumption is that in the inertial range $E(k)$ is determined by $\bar{\epsilon}$ and k and does not depend upon the phenomena at long or short length scales. This leads to

$$E(k) = C_K \bar{\epsilon}^{2/3} k^{-5/3}, \tag{1.14}$$

where C_K is a universal number found to be somewhere between 1.5 and 2.0 in many experiments [13, 14]. If we use dimensional analysis to get the characteristic decay rate Γ in this stationary state, we find

$$\Gamma(k) = \Gamma_0\, \bar{\epsilon}^{1/3} k^{2/3} \tag{1.15}$$

where Γ_0 is another universal constant. It is clear from (1.13) and (1.14), that in three dimensions

$$C(k) = \int C(k, \omega) \mathrm{d}\omega \sim k^{-11/3} \tag{1.16}$$

In general, in D-dimensions, $C(k) \sim k^{-(D+(2/3))}$. To probe the energy transfer rate, let us write the Navier–Stokes equation ((1.7)), in momentum space as

$$u_i(k) + \nu k^2 u_i = \sum_{\mathbf{p}} M_{ijl}(\mathbf{k}) u_j(\mathbf{p}) u_l(\mathbf{k} - \mathbf{p}) + f_i(\mathbf{k}) \tag{1.17}$$

where

$$M_{ijl}(\mathbf{k}) = k_j P_{il}(\mathbf{k}) + k_l P_{ij}(\mathbf{k})$$

and

$$P_{ij}(k) = \delta_{ij} - \frac{k_i k_j}{k^2}.$$

From (1.17), we have

$$\frac{\partial}{\partial t} \sum_i u_i^2(\mathbf{k}) + \nu k^2 \sum u_i^2(k) - f_i u_i(-\mathbf{k}) = \sum_{\mathbf{p}} M_{ijl}(\mathbf{k}) u_i(-\mathbf{k}) u_j(\mathbf{p}) u_l(\mathbf{k} - \mathbf{p}) \tag{1.18}$$

If we integrate over all k in (1.18), then

$$\int \mathrm{d}^D k \sum_{\mathbf{p}} M_{ijl}(-\mathbf{k}) \langle u_i(\mathbf{k}) u_j(\mathbf{p}) u_l(\mathbf{k} - \mathbf{p}) \rangle = 0$$

in the stationary state. We define the quantity

$$\prod(k) = \int_0^k \mathrm{d}^D k' \sum_{\mathbf{p}} M_{ijl}(-\mathbf{k}) \langle u_i(\mathbf{k}') u_j(\mathbf{p}) u_l(\mathbf{k} - \mathbf{p}) \rangle, \tag{1.19}$$

which now stands for the rate at which energy is flowing from wavenumbers below k to wavenumbers above k. Kolmogorov picture asserts that $\Pi(k)$ is independent of k.

We now turn to the randomly forced model. To find out how the relaxation rate $\Gamma(k)$ scales with k, we use the effective viscosity argument of Heisenberg. In the inertial range, where the non-linear term in Navier–Stokes equation is dominant, it is this non-linear term which gives rise to an effective scale dependent viscosity. Noting that the velocity field can be written as

$$u_i(\mathbf{k},t) = \int \mathrm{d}^D p \int \mathrm{d}t' G_{in}(k, t-t') M_{njl} u_j(\mathbf{p},t') u_l(\mathbf{k}-\mathbf{p},t') \tag{1.20}$$

the averaged effect of the non-linear term is clearly expressible as

$$\int \mathrm{d}^d p \mathrm{d}t' M_{ijl}(\mathbf{k}) M_{mrs}(\mathbf{p}) u_s(\mathbf{k}) G_{jm}(\mathbf{p}, t-t') \langle u_r(\mathbf{k}-\mathbf{p},t) u_l(-\mathbf{k}-\mathbf{p},t')\rangle.$$

The Greens function $G_{jn}(\mathbf{p}, t-t')$ can be written as

$$P_{jn}\,\mathrm{e}^{-\sum(\mathbf{p})(t-t')},$$

where $\sum(p)$ is the effective relaxation rate or the self energy in technical terms. The above expression then acquires the form

$$\int \mathrm{d}^D p\, \mathrm{d}t' M_{ijl}(\mathbf{k}) M_{mrs}(\mathbf{p}) P_{jm}(\mathbf{p}) P_{rl}(\mathbf{k}-\mathbf{p})$$
$$\times\, \mathrm{e}^{-\sum(p)(t-t')} C(\mathbf{k}-\mathbf{p}, t-t') u_s(\mathbf{k}, t')$$

where $C(k-p, t-t')$ is the time Fourier transform of $C(\mathbf{k}-\mathbf{p}, \omega)$. Now $C(q,\omega)$, has the form $q^{-D+4-y}[\omega^2 + \sum^2(q)]^{-1}$ and thus the last expression reduces to

$$\int T_{is}(\mathbf{k}, t-t') u_s(\mathbf{k}, t') \mathrm{d}t'$$

where the normalized trace of $T_{is}(\mathbf{k}, t-t')$ is the effective relaxation rate, whose Fourier transform yields the frequency dependent rate $\sum(k,\omega)$ as,

$$\sum(k,\omega) = \int \mathrm{d}^D p \mathrm{d}\omega' M_{ijl}(\mathbf{k}) M_{mri}(\mathbf{p}) P_{jm}(\mathbf{p}) P_{rl}(\mathbf{k}-\mathbf{p})$$
$$\times\, G(\mathbf{p},\omega') C(\mathbf{k}-\mathbf{p}, \omega-\omega') \tag{1.21}$$

Noting that $\int C(k,\omega)\mathrm{d}\omega \sim [\sum(k)]^{-1} k^{-D+4-y}$, it is easy to see that power counting consistency of (1.21) gives

$$\sum(k) \propto k^{2-(y/3)}. \tag{1.22}$$

Turning to (1.19) using (1.20) and following an identical line of argument that led to (1.22), we find that $\Pi(k) \sim k^{4-y}$. Kolmogorov asserts that $\Pi(k) = \bar{\epsilon}$ and hence one requires $y = 4$ for the Kolmogorov picture of turbulence [15–18]. That raises the central question is the theory well behaved at $y = 4$?

2. Technicalities

Kolmogorov scaling and the randomly stirred Navier–Stokes equation would be completely compatible if (1.21) would yield in addition to the power counting consistency as demonstrated towards the end of the last section, a finite number for universal amplitude ratios. With vertex corrections not yielding any new power laws because of Galilean invariance, the problem of turbulence would reduce to calculating the universal numbers in a loop ordered perturbation scheme. We now demonstrate that there are technical problems with this. To exhibit this, we explore the integrand on the rhs of (1.21). In particular, we probe the situation where the external momentum k is carried almost entirely by the propagator G and the momentum associated with the correlator C is very small. The zero frequency self energy can then be written as (suppressing projection operators and indices).

$$\sum(k) \simeq \int \mathrm{d}^D p \mathrm{d}\omega' M(\mathbf{k}) M(\mathbf{k}) G(k,\omega') C(p,-\omega'). \tag{2.1}$$

Now, $G^{-1}(k,\omega') = -i\omega' + \sum(k,\omega') \simeq \sum(k)$ as we cover the frequency range where ω' matches $p^{2/3}$ and hence is small compared to $\sum(k)$. Equation (3.1) thus becomes (for $y = 4$)

$$\begin{aligned}\sum(k) &\simeq \frac{M^2(k)}{\sum(k)} \int \frac{\mathrm{d}^D p}{(2\pi)^D} \frac{\mathrm{d}\omega}{2\pi} C(p,\omega) \\ &\simeq \frac{M^2(k)}{\sum(k)} \frac{\mathrm{d}^D p}{(2\pi)^D} \frac{1}{p^{D+2/3}}.\end{aligned} \tag{2.2}$$

This integral diverges due to the zero-p range contribution and needs to be cut off at some low momentum k_0. The scaling solution is consequently no longer valid.

The above divergence comes from the very strong dynamical coupling between the eddies (fourier modes) with short wavelength and the eddies with long wavelength. This effect is spurious. The expected role of the large eddies is to simply transport the much smaller ones and not to have a strong dynamical coupling. This spurious effect was attributed to the use of the Eulerian picture and Kraichnan tried to remedy it by going into the Lagrange description. This removes the problem in principle but actual calculations are virtually impossible. A different approach was tried by Yakhot and Orszag who used the standard renormalization group technique to circumvent the problem. The idea was that one splits the velocity field into two parts—one with high momentum Fourier modes and the other with low momentum Fourier modes, integrates out the high momentum components and studies the effect of that on the low momentum ones. This leads to a new viscosity and after the usual rescalings of space, time and the velocity field a flow equation for the viscosity emerges. The fixed point of the flow corresponds to a scale dependent viscosity. By construction, this yields long wavelength, low frequency properties as stressed by Forster *et al.* In the limit of extremely high Reynolds number, one is not particularly worried by this limitation. The process of integrating over the high momentum modes bypasses the difficulty that we encountered with the low-p divergence. We sketch the steps in the following:

(i) Split the velocity field as

$$u_i(\mathbf{x}) = u_i^{<}(\mathbf{x}) + u_i^{>}(\mathbf{x})$$

where

$$u_i^<(\mathbf{x}) = \sum_{p<\frac{\Lambda}{b}} u_i(\mathbf{p})\mathrm{e}^{i\mathbf{p}\cdot\mathbf{x}}$$

and

$$u_i^>(x) = \sum_{\frac{\Lambda}{b}<p<\Lambda} u_i(\mathbf{p})\mathrm{e}^{i\mathbf{p}\cdot\mathbf{x}}$$

(ii) Find the equations satisfied by $u_i^<$ and $u_i^>$.
(iii) Integrate perturbatively the equation for $u_i^>$
(iv) Insert the solution for $u_i^>$ into the equation for $u_i^<$ and find the equation of motion for $u_i^<$. The linear part gives $\Delta\nu$, the change in kinematic viscosity.
(v) Rescale space and length and the velocity field, so that the new viscosity is given by

$$\nu' = \nu b^{z-2} - 8b^{4-y-2z}F_0k^{-2}\int_{p=\frac{\Lambda}{b}}^{p=\Lambda}\frac{\mathrm{d}^Dp}{(2\pi)}D\int\frac{\mathrm{d}\omega}{2\pi}M_{\rho\beta\gamma}(\mathbf{k})$$
$$\times M_{\sigma\delta\gamma}(\mathbf{k}-\mathbf{p})P_{\beta\delta}(\mathbf{k}-\mathbf{p})P_{\sigma\rho}(\mathbf{p})G(\mathbf{k}-\mathbf{p}-\omega)|G(\mathbf{p},\omega)|^2p^{4-D-y}. \qquad (2.3)$$

Requiring that the two terms on the rhs have the same dimensions ensures

$$z = 2 - \frac{y}{3} \qquad (2.4)$$

and hence

$$\nu(k) = \nu k^{-y/3} \qquad (2.5)$$

as in (1.22). Evaluating the integral in (2.3) in the limit of $k \to 0$ gives

$$\nu' = \nu\left[1 + \ln b\left(\frac{D_0}{2\nu^3}\frac{D-1}{D+2}\frac{S_D}{(2\pi)^D} - \frac{y}{3}\right) + \cdots\right. \qquad (2.6)$$

where S_D is the surface area of a D-dimensional sphere. The fixed point condition yields

$$\frac{\nu^3}{D_0} = \frac{3}{2y}\frac{S_D}{(2\pi)^D}\frac{D-1}{D+2} \qquad (2.7)$$

From (1.19), one can determine the value of D_0^2/ν^3 and thus D_0/ν which sets the scale for the Kolmogorov spectrum is obtained. For $y = 4$ and $D = 3$, one does find a value very close to the experimental results and similar success in calculating other universal numbers certainly indicates that this is a successful program.

The primary difficulties of this approach were pointed out by De Dominicis and Martin [10] at the time they introduced the model at the Kolmogorov limit i.e. $y = 4$, there are an infinite number of marginal operators and the self consistent perturbation theory has an infrared divergence at $y = 3$. In the last few years, both these effects have been investigated in a different context—the problem of growth by deposition of atoms on a substrate [19–22]. The existence of an infinite number of marginal operators and infrared divergences seem to change the roughening and the dynamical exponents of the problem

[23–26]. Consequently, in spite of impressive numerical agreement with experiments, the renormalization group approach is somewhat suspect.

We now note that the infrared divergence that started the problem in the first place is caused by the strong dynamical coupling between small scale (large k-Fourier components) energy dissipating eddies and the large scale (small k-Fourier components) energy containing eddies. For the validity of Kolmogorov scaling, the small eddies should simply be advected by the large eddies without any dynamical coupling. Hence Kolmogorov scaling requires a screening [27, 28] of the interaction between the large and small Fourier coefficients. This is provided by viscoelasticity [29]. Turning to (1.21), we note that the effective self energy $\sum(k,\omega)$ is strongly frequency-dependent for frequencies greater than $\sum(k,0)$. If we write the k-dependence of the relaxation rate as

$$\Gamma(k) = \Gamma_0 k^n \tag{2.8}$$

and that of the equal time correlation function as

$$C(k) = c_0 k^{-m} \tag{2.9}$$

then dynamic scaling yields

$$\sum(k,0) = \Gamma_0 k^n \tag{2.10a}$$

and

$$\sum(k,\omega) = k^2(-i\omega)^{(2-n)/n} \quad \text{for } \omega \gg \Gamma(k). \tag{2.10b}$$

The corresponding correlation function is

$$C(k,\omega) = \frac{C(k)}{\sum(k)}\left[\frac{1}{1-i\omega/\sum(k,\omega)} + \frac{1}{1+i\omega/\sum^{+}(k,\omega)}\right] \tag{2.11}$$

for frequencies ω such that $\omega\tau_s \gg 1$, where τ_s is the sweeping time. Light scattering from the randomly stirred fluid under the right conditions [28] should be able to establish the validity of (2.11).

We now introduce the form of $C(k,\omega)$ given above and using a Lorentzian $G(k,\omega)$ i.e., $G(k,\omega) = -i\omega + \sum(k)$, arrive at (see eq. (2.1))

$$\sum(k) \sim \int \frac{\mathrm{d}^D p}{(2\pi)^D}\frac{C(p)}{\sum(p)}\left[1 + \frac{\sum(\mathbf{k}-\mathbf{p})}{\sum(p,\omega=-i\sum(\mathbf{k}-\mathbf{p}))}\right]^{-1}. \tag{2.12}$$

If we now explore the region of the integrand where $p \to 0$, then

$$\begin{aligned}\sum(k) &\sim \int \frac{\mathrm{d}^D p}{(2\pi)^D}\frac{C(p)}{\sum(p)}\left[1 + \frac{\sum(k)}{p^2[\sum(k)]^{(n-2/n)}}\right]^{-1} \\ &\sim \int \frac{\mathrm{d}^D p}{(2\pi)^D}\frac{C(p)}{\sum(p)}\,p^2 \\ &\sim \int \frac{\mathrm{d}^D p}{(2\pi)^D}C(p)\,p^{2-n}\end{aligned} \tag{2.13}$$

The infrared divergence now occurs for $m+n \geq D+2$. For the Kolmogorov case of $m=11/3$ and $n=2/3$ in $D=3$ one is safe.

The above manipulations bear some similarity (in the final results) to the quasi-Lagrangian [30] approach advocated by Belinicher and L'vov [31, 32]. This has recently been taken up again by L'vov and Procaccia [33] who work entirely in coordinate space to show that the Kolmogorov spectrum does lead to a finite theory in the perturbative analysis of the forced Navier–Stokes equation.

In all our considerations so for, dimensionality plays a very minor role. The exponent z is independent of dimension while the exponent n carries a trivial dependence on dimensionality. This implies that one may possibly consider a one dimensional randomly stirred model and pick up the features that we have talked about so far. In one dimension it is convenient to consider Burger's equation which is our Navier–Stokes equation without the pressure term. For the incompressible flows that we were dealing with, the pressure term simpy modified the coupling constant associated with the nonlinear term and hence did not cause any qualitative change. Dropping it and considering the randomly stirred Burgens equation is one dimension is not going to cause any qualitative changes in the central problem. Thus

$$\dot{u}+u\frac{\mathrm{d}}{\mathrm{d}x}u=\nu\nabla^2 u+f \tag{2.14}$$

with

$$\langle f(x,t)f(x',t')\rangle = D_0 k^{-1+y-4}\delta(x-x')\delta(t-t') \tag{2.15}$$

is expected to capture the essential features of fully developed Kolmogorov turbulence for $y=4$, there is an additional complication. Without the drive f, Burgers equation is known to have coherent structures in the form of shock waves [34]. To see Kolmogorov scaling, the random drive has to overcome the shocks. This idea that (2.14) contains the basic ingredients for understanding fully developed turbulence has recently been supported very stronglly by Cheklov and Yakhot [35], Polyakov [36] and Gurarie and Migdal [37].

The thrust of our discussion so far has been that Kolmogorov scaling is exact. Our calculations have been in the inertial range where the effective relaxation rate $\sum(k)$ dominates the molecular relaxation rate νk^2. This means that we are working with $k \ll k_D$ where k_D is the wavenumber formed from the rate of dissipation ϵ and the kinematic viscosity ν. Clearly

$$k_D=\left(\frac{\epsilon}{\nu^3}\right)^{1/4} \tag{2.16}$$

and in the inertial range $k_D L \gg 1$, where L is the system size. With $\epsilon \sim v^3/L$ (v is a mean velocity)

$$k_D L \sim (Re)^{3/4} \tag{2.17}$$

Where $Re = vL/\nu$ is the Reynold's number and the self consistent approach that we have adopted (and which works!) requires $Re \to \infty$. Thus, in the infinite Reynold's number limit, the results for the relexation rate and the two point function $\langle v(k,\omega)v(k',\omega')\rangle$ (alternatively $\langle [v(x+r,t)-v(x,t)]^2\rangle$) seem to be exact.

We now have to face the fact that a whole class of experiments yield results for higher order structure factors [38–41] that are at variance with the Kolmogorov expectations. If one investigates the structure factor $\langle |v(x+r,t) - v(x,t)|^P \rangle$, then the exponent specifying the r-dependence does not depend linearly on p. On the other hand assuming that once the two point function has been shown to be finite, there would be no new divergences in the higher order correlations and hence the exponent for the pth-order structure factor should be a linear function of p. The correlation function $\langle \epsilon(x+r)\epsilon(x) \rangle$ in particular is of great interest ($\epsilon(x)$ is the dissipation rate at the point x, averaged over a local area) since in the Kolmogorov picture it has no r-dependance. Various experiments reveal $\langle \epsilon(x+r)\epsilon(x) \rangle \sim r^{-\mu}$ with μ a small number (about 0.20). This phenomenon is known as intermittency and μ is often called the intermittency exponent. We will return to this issue in § 3.

The Kolmogorov picture in two dimensions is special because there are two cascades to contend with—energy and enstrophy. Enstrophy is defined as $(\nabla \times \mathrm{v})^2$. The direction of the cascades can be determined by an argument due to Kraichnan [42]. With two conserved quantities,

$$E = \frac{1}{2}\int v^2(x)\mathrm{d}^2x = \frac{1}{2}\sum_k |v(k)|^2 = \int_0^\infty E(k)\mathrm{d}k$$

and

$$N = \frac{1}{2}\int (\nabla \times \mathrm{v})^2 \mathrm{d}^2x = \frac{1}{2}\sum_k k^2|v(k)|^2 = \int W(k)\mathrm{d}k$$

where $E(k)$ and $W(k)$ are the energy and enstrophy spectrum respectively. The canonical probability distribution for fluctuating E and N is clearly

$$P \sim \mathrm{e}^{-\beta E - \mu N} = \mathrm{e}^{-\frac{1}{2}\sum(\beta + \alpha k^2)|v_k|^2}. \tag{2.18}$$

This leads to

$$\langle |v_k|^2 \rangle = (\beta + \alpha k^2)^{-1} \tag{2.19}$$

and consequently

$$E(k) = \frac{k}{\beta + \alpha k^2}$$

and

$$W(k) = \frac{k^3}{\beta + \alpha k^2}$$

The enstrophy spectrum $W(k) \sim k$ for high wave number and hence is clearly concentrated towards high wave number side. A spectrum with $W(k) \sim k^x$, $x < 1$, will be out of equilibrium and will proceed towards equilibrium by cascading enstrophy from low to high wave numbers. Energy conservation would then demand an inverse cascade of energy. The enstrophy cascade is dissipated by molecular viscosity at high wave vectors, while the inverse energy cascade causes a condensation phenomenon at low wave numbers. The Kolmogorov argument holds for the inverse energy cascade and one ends p with $E(k) \sim k^{-5/3}$, while for the enstrophy cascade one has to set up the dimensional rgument once more, postulating that in the inertial range $E(K)$ is determined by the rate f injection of enstrophy (ϵ_s say) and the local wave number k. This leads to $E(k) \sim k^{-3}$. he infrared difficulties that we talked about in the beginning of the section will appear ere as well and a screening approximation will produce a self energy of the form given ı (2.13). For the energy cascade in $D = 2$, $m = 8/3$ and $n = 2/3$ and the inequality $\imath + n < D + 2$ is satisfied which makes the theory finite. For the enstrophy cascade on ıe other hand, $m = 4$ and $n = 0$ and we have $m + n = D + 2$. That means a genuine ogarithimic divergence and hence for the enstrophy cascade [44–46]

$$E(k) = c_0(\epsilon_s)^{2/3} k^{-3} (\ln k/k_0)^{-1/3}. \tag{2.20}$$

et another problem where there exists multiple cascades is the magnetohydrodynamic ırbulence [47–50]. The conserved quantities in the inviscid limit are the kinetic energy nd the magnetic energy. Consequently, there is a magnetic energy flux and a kinetic nergy flux and one can determine [51] the Kolmogorov constants etc., for this flow in a ıanner analogous to that for the pure fluid. The binary liquid is another example of a tuation with different fluxes [52]—an energy flux and a concentration flux and a enormalization group program for that has been carried out [53, 54] recently.

. Subtleties

his short section has to do with the question whether Kolmogorov scaling is exact for ıe forced Navier–Stokes equation or not. To set the stage, we begin with the earliest iscussion [54, 55] of this topic. Soon after the publication of Kolmogorov's work, it was ointed out by Landau that the theory could be internally inconsistent. The dissipation ate ϵ is found from Navier–Stokes equation to be expressible as

$$\epsilon = \nu \int \mathrm{d}^D r (\mathrm{v} \cdot \nabla^2 \mathrm{v})^2 = \sum_{\alpha,\beta} \frac{\nu}{2} \int \mathrm{d}^D r \left(\frac{\partial v_\alpha}{\partial x_\beta} + \frac{\partial v_\beta}{\partial x_\alpha} \right)^2. \tag{3.1}$$

eing an integral over a fluctuating quantity, namely the velocity field, the dissipation ate ϵ is expected to show fluctuations which are ignored in the Kolmogorov picture. It as not till two decades had passed that Kolmogorov and Obukhov returned to this issue nd took Landau's objection into account by assuming that the local dissipation rate ϵ_r at ıe point r (averaged over a small ball around r) had a log normal distribution i.e., $\ln\epsilon_r$ ad a Gaussian distribution.

Kolmogorov assumed that the width σ of the distribution had the form

$$\sigma^2 = \langle [\ln\epsilon_r - m]^2 \rangle = \begin{cases} A + 9\delta \ln L/r, & k_d r \gg 1 \\ A' + 9\delta \ln k_D L, & k_D r \ll 1 \end{cases} \tag{3.2}$$

Vhere $m = \langle \ln\epsilon_r \rangle$ and δ is an universal number. The constants A and A' are determined by ıe large scale structure of the flow. The log normal distribution of ϵ_r requires that the robability distribution $P_r(a)$ for $\epsilon_r = a$ is given by

$$P_r(a) = \frac{1}{(2\pi\sigma^2 a^2)^{1/2}} \mathrm{e}^{-(\ln a - m)^2/2\sigma^2}. \tag{3.3}$$

This distribution leads to

$$\langle \epsilon_r^p \rangle = e^{pm+\frac{1}{2}p^2\sigma^2} \tag{3.4}$$

and subsequently

$$\langle |\mathbf{v}(\mathbf{x}+\mathbf{r},t) - \mathbf{v}(\mathbf{x},t)|^p \rangle \sim (r\bar{\epsilon})^{p/3} \left(\frac{L}{r}\right)^{(\delta/2)p(p-3)} \tag{3.5}$$

The intermittency exponent defined in the previous section is obtained from $p = 6$ and found to be $\mu = 9\delta$.

Experimental results on the high order structure factors do not quite agree with the p-dependence of the exponent shown in (3.5), but they certainly follow the qualitative feature that the exponents are not a linear function of p. The deviation from linearity prompts a multifractal description, which in turn inspires a dynamical system description of turbulence in the form of shell models. We shall not discuss these issues here they will be addressed in detail by Pandit *et al* [56] in this volume. We will confine our discussion to the randomly stirred model.

The discussion of the previous section leads us to believe that the scaling of the two point function is exact in the high Reynold's number limit (finite Reynolds number corrections [57,58] have been found which provide correction-to-scaling and hence an appearance of deviation from Kolmogorov scaling). However, the case of higher order correlation functions had not been discussed till almost the mid-nineties. The natural quantity to focus on first is the exponent μ and a lowest order calculation of the corresponding correlation function showed [59] a logarithmic behaviour i.e., $\langle \epsilon(\mathbf{x}+\mathbf{r})\epsilon(\mathbf{x}) \rangle \sim \bar{\delta} \ln(L/r)$. For small μ, the $r^{-\mu}$ behaviour can be expanded to give $\mu \ln(1/r)$ and $\bar{\delta}$ can be identified with μ. This yields a reasonable value of μ. The work of Lebedev and L'vov [60] and subsequent calculations [61, 62] examine the structure of the relevant correlation function to all orders and conclude that a resummation can lead to a power law. However, they do not calculate the exponent explicitly.

Recent developments [35–37] very strongly support the idea that the higher order structure factors in the randomly stirred fluid do show deviations from pure Kolmogorov scaling. These have to do with the one dimensional Burgers equation-turbulence discussed before. Numerical work [35] does clearly show a non-zero μ and analytical work from two different directions—operator products expansion [36] and WKB type analysis—reveals a non-Gaussian probability distribution [63].

Acknowledgements

The author is thankful to Dr Rahul Pandit for bringing references [36] and [37] to his attention.

References

[1] H Tennekes and Lumley, *A first course in turbulence* (MIT Press Cambridge, Massachusetts, 1972)

[2] D C Leslie, *Developments in the theory of turbulence* (Clarendon Press, Oxford, 1972)

[3] S A Orszag, *Les Houches lectures on fluid dynamics* (Gordon and Breach, London) 1973
[4] A S Monin and A M Yaglom, *Statistical fluid mechanics* (MIT Press, 1971) Vol. 2
[5] M Lesieur, *Turbulence* (M Ninjhuis, Amsterdam, 1988)
[6] W D McComb, *The physics of fluid turbulence* (Clarendon Press, Oxford, 1990)
[7] L Sirovich ed, *New perspectives in turbulence* (Springer–Verlag, Berlin, 1990)
[8] E Ott, *Chaos in dynamical systems* (Cambridge University Press, 1993)
[9] D Forster, D Nelson and M Stephen, *Phys. Rev.* **A16** 732 (1976)
[10] C De Dominicis and P C Martin, *Phys. Rev.* **A19**, 419 (1979)
[11] P C Martin and C De Dominics, *Prog. Theor. Phys.* (Suppl.) **64**, 108 (1978)
[12] A N Kolmogorov, *C R (Dokl) Acad. Sci. URSS* **30**, 301 (1941)
[13] H L Grant, R W Stewart and A Moillet, *J. Fluid Mech.* **12**, 241 (1962), **13**, 237 (1962)
[14] C W Van Atta and J Park, in *Statistical models and turbulence* (Springer-Verlag, New York, 1972)
[15] V Yakhot and S A Orszag, *Phys. Rev. Lett.* **57**, 1722 (1986); *J. Sci. Comput.* **1**, 1 (1986)
[16] J K Bhattacharjee, *Phys. Rev.* **A40**, 6374 (1989)
[17] D Ronis, *Phys. Rev.* **A36**, 3322 (1987)
[18] E V Teodorovich, *Sov. Phys. Dokl.* **33**, 247 (1988); *Prikl. Matem. Mekhan.* USSR **53**, 340 (1989), *Sov. Phys. JETP* **75**, 472 (1992)
[19] J K Bhattacharjee, *J. Phys.* **A21**, L551 (1988)
[20] M Kardar, G Parisi and Y Zhang, *Phys. Rev. Lett.* **56**, 889 (1986)
[21] D W Wolf and J Villain, *Europhys. Lett.* **13**, 389 (1991)
[22] Z W Lai and S Das Sharma, *Phys. Rev. Lett.* **66**, 2348 (1991)
[23] J Krug, *Phys. Rev. Lett.* **72**, 2907 (1994)
[24] S Das Sharma, S V Ghaisas and J M Kim, *Phys. Rev.* **E49**, 122 (1994)
[25] J K Bhattacharjee, S Das Sharma and R Kotlyar, *Phys. Rev.* **E53**, R1313 (1996)
[26] C Das Gupta, S Das Sharma and J M Kim, preprint (1996)
[27] J K Bhattacharjee, *Phys. Fluids* **A3**, 879 (1991); *Mod. Phys. lett.* **B7**, 881 (1993)
[28] J K Bhattacharjee, *J. Phys.* **A27**, L347 (1994)
[29] S C Crow, *J. Fluid Mech.* **33**, 1 (1968)
[30] H Horner and R Lipowsky, *Z. Phys.* **B33**, 233 (1979)
[31] V I Belinicher and V S L'vov, *Sov. Phys. JETP* **66**, 303 (1987)
[32] V S L'vov, *Phys. Rep.* **207**, 1 (1991)
[33] V S L'vov and I Procaccia, *Phys. Rev.* **E52**, 3840, 3858 (1995)
[34] P Saffman, in *Topics in Nonlinear Physics* edited by N J Zabusky (Springer Verlag, New York, 1968)
[35] A Chekhlov and V Yakhot, *Phys. Rev.* **E51**, R2739 (1995)
[36] A M Polyakov, *Phys. Rev.* **E52**, 6183 (1995)
[37] V Gurarie and A Migdal, preprint (1996)
[38] F Anselmet, Y Gagne, E J Hopfinger and R A Antonia, *J. Fluid mech.* **140**, 63 (1984)
[39] C Meneveau and K R Sreenivasan, *Phys. Rev. Lett.* **59**, 1424 (1987)
[40] C Meneveau and K R Sreenivasan, *Nucl. Phys.* (Proc. Suppl.) **B2**, 49 (1987)
[41] K R Sreenivasan, *Ann. Rev. Fluid Mech.* **23**, 539 (1991)
[42] R H Kraichnan, *Phys. Fluids* **10**, 1417 (1967)
[43] G K Batchelor, *Phys. Fluids Suppl.* **12**, II-233 (1969)
[44] R H Kraichnan, *J. Fluid Mech.* **47**, 525 (1971)
[45] P Olla, *Phys. Rev. Lett.* **67**, 2464 (1991)
[46] M K Nandy and J K Bhattacharjee, *Int. J. Mod. Phys.* **B9**, 1081 (1995)
[47] R H Kraichnan, *Phys. Fluids* **8**, 1385 (1965)
[48] M Dobrowolny, A Mangeney and P Veltri, *Phys. Rev. Lett.* **45**, 144 (1980)
[49] W H Matthaeus and Y Zhon, *Phys. Fluids* **31**, 3634 (1988)
[50] S J Camargo and H Tasso, *Phys. Fluids* **B4**, 1199 (1992)
[51] M K Verma and J K Bhattacharya, *Europhys. Lett.* **31**, 195 (1995)
[52] M K Nandy, *Topics in homogeneous isotropic turbulence* Ph.D. Thesis IIT Kanpur, (1995).
[53] M K Nandy and J K Bhattacharjee, submitted (1996)
[54] A M Obukhov, *J. Fluid Mech.* **13**, 77 (1962)

55] A N Kolmogorov, *J. Fluid Mech.* **13**, 82 (1962)
56] S K Dhar, A Sain, A Pande and R Pandit, *Pramana – J. Phys.* **48**, 325 (1997)
57] S Grossman, D Lohse, V L'vov and I Procaccia, *Phys. Rev. Lett.* **73**, 432 (1994)
58] V S L'vov and I Procaccia, *Phys. Rev. Lett.* **74**, 2690 (1994)
59] A Das and J K Bhattacharjee, *Europhys. Lett.* **26**, 527 (1994)
60] V V Lebedev and V S L'vov, *JETP Lett.* **59**, 577 (1994)
61] V S L'vov, I Procaccia and A L Fairhall, *Phys. Rev.* **E50**, 4684 (1994)
62] V S L'vov and V V Lebedev, *Phys. Rev.* **E47**, 1794 (1993)
63] R H Kraichnan, *Phys. Rev. Lett.* **65**, 575 (1990)

Chaos in the hydrogen atom interacting with external fields

K GANESAN and R GĘBAROWSKI
Department of Applied Mathematics and Theoretical Physics, The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland, United Kingdom

Abstract. In this review we discuss the chaotic dynamics (both classical and quantal aspects) of a simple atomic system, namely hydrogen atom interacting with time independent and time dependent external fields. These include: i) static electric field, ii) static magnetic field, iii) combined electric and magnetic fields, in parallel and perpendicular configuration, iv) instantaneous and generalized van der Waals field, v) mass anisotropy and vi) linearly and circularly polarized microwave fields.

Keywords. Classical and quantum chaos; hydrogen atom in external fields; Stark and Zeeman effect; van der Waals field; linearly and circularly polarized microwave; ionization; diffusion; Rydberg states.

PACS Nos 05.45; 31.15; 32.30; 32.60; 32.80

1. Introduction

Most of the naturally occurring phenomena belonging to various disciplines of science are generally governed by nonlinear differential equations, but for the sake of mathematical tractability and simplicity they have been approximated to linear forms. In recent years, with the advent of powerful computers and advanced numerical techniques, these nonlinear differential equations can be solved to a greater extent. Moreover analysing the long time behaviour of such systems is also feasible now.

In addition to the above, recently many advanced analytical techniques have been devised to solve some of these nonlinear differential equations. In fact a large class of nonlinear partial differential equations are shown to exhibit so-called soliton solutions [1]. Even though these systems are governed by infinite dimensional equations, their solutions exhibit, coherent and regular structures. On the other extreme there are systems governed by very simple nonlinear (differential) equations whose (numerical) solutions exhibit complicated structures. This class of systems whose solutions are very sensitive to initial conditions are generally called chaotic systems [2].

Generally the dynamics of macroscopic systems follows the principles of classical mechanics while microscopic systems (atomic and sub atomic systems) obey the principles of quantum mechanics. The definition and the meaning of chaos in classical systems is well understood now. In fact there are many numerical techniques (both qualitative and quantitative) to characterize chaos in classical systems. On the other hand the underlying equation of motion in quantum mechanics, namely the Schrödinger equation, is linear in nature. Moreover the finiteness of Planck's constant impose problems in carrying over the tools used for defining chaos in classical mechanics. Hence characterizing chaos in quantum systems is still a daunting task. However significant breakthroughs have been achieved by probing the appropriate quantum mechanical quantities (mainly eigenvalues and eigenfunctions). In fact some fingerprints and signatures of chaos in quantum systems are well known [3, 4]. In this review we concentrate on the manifestations of chaos in atomic systems which are not only suitable for both classical and quantal treatments but also amenable to real time laboratory experiments.

By classical chaos in atomic systems we mean that the atomic electron moves in a seemingly erratic fashion in the phase space (although governed by deterministic equations) and as a result of it, explores most of the available phase space if allowed to evolve for sufficiently long time. Even if the electron starts its journey from an infinitesimally different point in the phase space, its dynamics will follow an entirely different path. It has been shown that the paths separate exponentially and the degree of their separation is characterized by the so-called Lyapunov exponent [2]. This exponential deviation of nearby trajectories is characteristic of many nonlinear and non-integrable (i.e. which do not posses sufficient number of independent constants of motion) systems.

Energy level clustering is the hallmark of quantal systems whose classical counterparts exhibit regular motion. That means there is a large probability for small spacing and the spacing between adjacent levels follow the Poisson distribution. Level repulsion is the predominant fingerprint of chaos in quantum systems. As a result of this, the spacing between the adjacent energy levels follows the Wigner distribution. Moreover, if we probe different regions of the parameter space, one may be able to notice many avoided level crossings. In fact this conjecture has been successfully verified in many model calculations [5]. In this review we plot the energy levels of different atomic systems for a range of parametric values.

The hydrogen atom is the simplest atomic system. Pure hydrogen atom is of not much interest as both its classical and quantal behaviour are well known. But hydrogen atom in an external electromagnetic field turns out to be one of the best testing grounds for understanding the concepts of chaos [2–4]. In § 2 of this review we consider the chaotic dynamics of hydrogen atom in time independent external fields which are accessible for laboratory experiments such as static electric or/and magnetic fields. Section 3 deals with the chaotic dynamics of hydrogen atom in other interesting external fields such as van der Waals field and mass anisotropy. In § 4 we discuss the chaotic dynamics of the hydrogen atom in time dependent external electromagnetic fields. In particular we concentrate on linearly and circularly polarized microwave fields. In § 5, we enumerate our conclusions.

2. Hydrogen atom in static electric or/and magnetic fields

We assume that the nucleus has an infinite mass and is at rest, and neglect the spin of the electron for simplicity. Relativistic effects are negligible and further we always consider the vanishing magnetic quantum number case wherever it is a good quantum number. All Hamiltonians are written in dimensionless atomic units ($m_e = |e| = |\hbar| = 1$) unless otherwise stated.

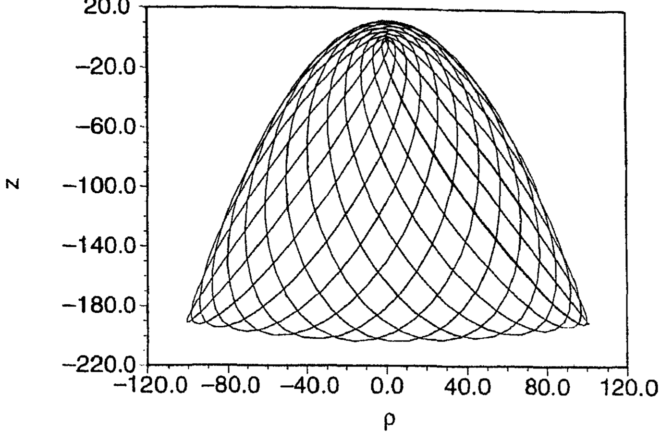


Figure 1. The (regular) trajectory in the ρ-z plane of a hydrogen atom in an electric field associated with the Hamiltonian H_{ef} for $F = 9950\,\mathrm{V/cm}$ and $n = 10$ manifold of the field free case.

2.1 *Hydrogen atom in an electric field (Stark effect)*

To begin with, let us consider the well known and the simplest external field problem, namely the hydrogen atom in a static electric field. The Hamiltonian of the hydrogen atom in an uniform static electric field applied parallel to the z-axis is given as [6, 7]

$$H_{\mathrm{ef}} = \frac{p^2}{2} - \frac{1}{r} + Fz. \tag{2.1}$$

Here r and p are the position and momentum coordinates expressed in three dimensional Cartesian coordinate system. F represents the strength of the external electric field usually measured in the units of V/cm. In some sense this is one of the least complicated problems: it is simple because it is separable (and hence integrable) in many coordinate systems. That means one can find the required number of independent constants of motion, so its classical behaviour is regular for arbitrary field strengths F [3]. In figure 1, we have shown a representative classical trajectory (in cylindrical coordinates which are defined as $\rho = \sqrt{x^2 + y^2}$, $\phi = \tan^{-1}(y/x)$, $z = z$) for a (high) field strength of F=9950 V/cm and an energy value which corresponds to $n = 10$ manifold of the field free case. From figure 1 it is evident that the electron follows a well defined regular path.

The electric field problem at the same time is complicated in a sense that any atomic state ionizes spontaneously in the presence of an uniform electric field. The ionization process exhibits a threshold behaviour which is related to the critical point (saddle point) of the underlying potential energy surface. The energy of the saddle point is given as $E_{\mathrm{sp}} = E_i - 2\sqrt{F}$ where E_i is the ionization potential of the unperturbed atom (it is the amount of energy required to remove an electron out of atom — we choose $E_i = 0$) [8]. For energy less than E_{sp}, the excited electron remains essentially bound and the Stark states are quasi-stable with respect to the ionization, while for energy higher than E_{sp}, the atom ionizes. At this stage, the quasi-stable energy levels of the atomic

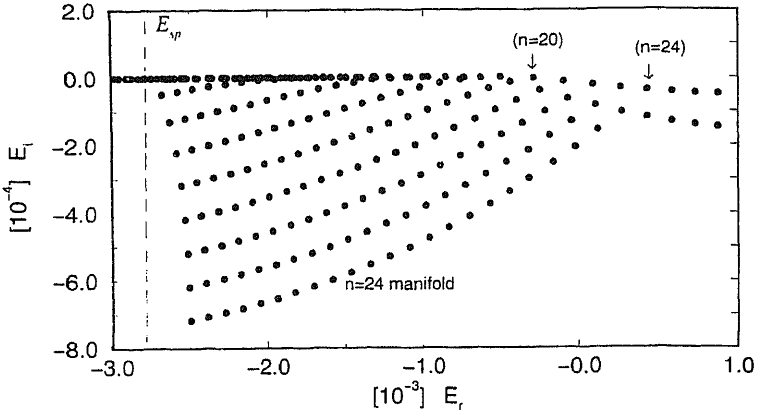


Figure 2. The regular arrangement of resonances (complex eigenvalues) of a hydrogen atom in an electric field associated with the Hamiltonian H_{ef} for $F = 9950\,\mathrm{V/cm}$ from just below the classical ionization limit E_{sp} up to positive energies.

states are superimposed on ionization continua. Now the atomic states are associated with complex energy eigenvalues whose real part E_r gives the energy and imaginary part E_{im} gives the half-width of the decaying states [8, 9].

In figure 2 we have shown the resonances (the real and imaginary parts of the complex eigenvalues) of the hydrogen atom in an electric field of strength $F = 9950$ V/cm from just below the classical ionization limit E_{sp} up to positive energies. Far from being randomly scattered around, they appear in lines spreading out in the lower half of the complex energy plane. Each such string is made up of resonances of one n-manifold [9].

Thus what we notice is that in the case of hydrogen atom in an uniform static electric field, the bounded classical trajectories are regular while the quantum resonances (above the classical saddle point) also follow a definite regular pattern for any arbitrary electric field strength. This might be due to the integrable nature of the problem.

2.2 *Hydrogen atom in a magnetic field (Zeeman effect)*

Unlike the electric fields, magnetic fields do not ionize atoms. They tend to compress the charge to the nucleus and increase the binding. Magnetic fields have played useful and important role in the history of atomic physics. In fact the Zeeman effect provided the first direct evidence of the electromagnetic origin of light and the Stern–Gerlach experiment proved the reality of spatial quantization. The Hamiltonian of the hydrogen atom in an uniform static magnetic field can be written as [3, 10]

$$H_{\mathrm{mf}} = \frac{p^2}{2} - \frac{1}{r} + \frac{B}{2}L_z + \frac{B^2}{8}(x^2 + y^2), \tag{2.2}$$

where L_z is the z-component of the angular momentum and B is a measure of the strength of the magnetic field usually measured in units of Tesla. For weak magnetic fields, the last term which is quadratic in B (usually called the diamagnetic term), is negligible. By treating the third term (the paramagnetic term) as a perturbation, we obtain the familiar result (available in many text books on quantum mechanics) [11]

$$E(n,l,m) = -\frac{1}{2n^2} + \frac{m}{2}B. \tag{2.3}$$

Here n, l and m are respectively the principal, orbital and magnetic quantum numbers. In this perturbative regime, all the three quantum numbers n, l and m are good, and therefore the linear Zeeman term, for a fixed magnetic quantum number 'm', just adds a static quantity, a constant shift. As L_z (which is quantized as $L_z = m\hbar$) is a constant of motion, the classical motion remains regular in this regime.

2.2.1 *Quadratic Zeeman effect*: If we include the quadratic term in magnetic field as well, many interesting features come into play [12, 13]. In fact one has to include these quadratic terms when the strength of the magnetic field is high. The leading terms in the average value of $\langle x^2 + y^2 \rangle$ are of the order of $B^2 n^4$, while for the Coulombic potential the gap between adjacent n-manifolds is of the order of n^{-3}. Hence we see that the diamagnetic term has greatest effect at high n values. In fact there are 3 regimes for this system, 1) low, 2) high and 3) intermediate fields [14]. We discuss the salient features of each of these in the following subsections.

Low field regime. In this regime the diamagnetic term is treated as a perturbation to the Coulombic potential. Here l-mixing (l is not a good quantum number: at moderate field strengths different l-states within a given n-manifold mix) occurs but there is no n-mixing. As a result, the classical motion remains predominantly regular in this regime, and for small but finite field strengths, there exist an adiabatic invariant (discussed later in detail for the Stark-quadratic Zeeman problem) which is helpful in identifying the so called rotator and vibrator states [15]. Actually the ratio of the diamagnetic and Coulombic terms, namely $B^2 n^7$ mainly determines the behaviour of the system. It has been shown qualitatively that the onset of n-mixing occurs when $B^2 n^7 \simeq \frac{16}{5}$, so for $n = 6$ manifold one needs to have a magnetic field of strength about 794 Tesla while for $n = 80$ manifold only 0.09 Telsa is sufficient to induce n-mixing. Therefore it is clear that for highly excited states (the Rydberg states), a small magnetic field is sufficient to induce n-mixing; energy levels belonging to different n-manifolds may interpenetrate [12], and the perturbative treatment is not valid.

High field regime. For very high magnetic fields, the diamagnetic potential dominates and the Coulomb potential may be neglected. This is the Landau regime [11], and the Hamiltonian simplifies to

$$H_{\text{hmf}} = \frac{p^2}{2} + \frac{B^2}{8}(x^2 + y^2). \tag{2.4}$$

If we neglect the free motion in the z direction, we get a two dimensional isotropic harmonic oscillator with oscillator energy $\hbar\omega = B, \omega = \frac{1}{2}\omega_c$ (ω_c is the cyclotron frequency). The quantum energy levels are given by

$$E_{\mathrm{Lan}} = (N+\tfrac{1}{2})B, \tag{2.5}$$

where $N = 0, 1, 2, \ldots$ labels the Landau levels. If we include the Coulombic terms perturbatively, then one can show that [16] the quantum energy levels are expressed as

$$E_{Nm} = E_m + NB, \quad E_m = (|m|+1)\frac{B}{2}, \tag{2.6}$$

where 'm' is the magnetic quantum number. The classical motion is regular in this regime.

Quasi-Landau resonances. In 1969, Garton and Tomkins [17] discovered a dramatic modulation in the absorption spectrum of barium in a field of 2.5 Tesla (hereafter abbreviated as T). The modulation pattern now called as quasi-Landau resonances, extends above the zero-field ionization limit of the atom and has a period close to $1.5\omega_c$ (ω_c is the cyclotron frequency). At higher energies, the period approaches ω_c, as might be expected for a free electron.

To understand these periods, we neglect the z-motion and treat the problem as 2 dimensional. Using the primitive WKB expression

$$\int_{\rho_1}^{\rho_2} \sqrt{2E - 2V(\rho)}\,\mathrm{d}\rho = (n+\tfrac{1}{2})\pi, \tag{2.7}$$

where $\rho = \sqrt{x^2+y^2}$, ρ_1 and ρ_2 are the classical turning points and $V(\rho)$ is the effective 2 dimensional potential given by

$$V(\rho) = -\frac{1}{\rho} + \frac{L_z^2}{2\rho^2} + \frac{B}{2}L_z + \frac{B^2}{8}\rho^2, \tag{2.8}$$

the separation between the levels can be found as [18]

$$\frac{\mathrm{d}E}{\mathrm{d}n} = \pi\left[\int_{\rho_1}^{\rho_2} \sqrt{2E - 2V(\rho)}\,\mathrm{d}\rho\right]^{-1}. \tag{2.9}$$

Integrating this equation between the turning points at $E = 0$, we get $\mathrm{d}E/\mathrm{d}n = 1.5\omega_c$. Numerical evaluation for higher energies shows that the spacing slowly diminishes. Several researchers have studied these quasi-Landau resonances experimentally and confirmed these findings [19].

Intermediate field regime. In this regime, the magnetic field strength becomes comparable to the Coulombic field strength. The problem loses both its spherical symmetry (due to the Coulombic term) and cylindrical symmetry (due to the quadratic Zeeman term). The problem is not only non-separable and non-integrable, but reaches a non-perturbative regime. The problem cannot be solved by any analytical method. One needs to resort to numerical methods.

In this regime, adjacent n-manifolds interpenetrate [20]. The levels repel strongly and the regularities of the energy spectrum are lost. In figure 3 we show a chaotic classical trajectory (in cylindrical coordinates) at a magnetic field strength of $B = 7$ T and for an

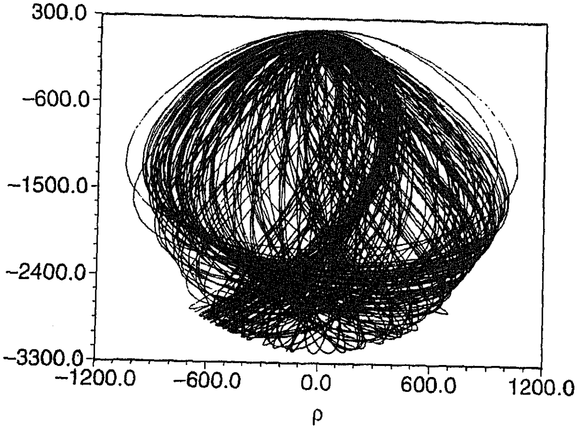


Figure 3. The (chaotic) trajectory in the ρ-z plane of a hydrogen atom in a magnetic field associated with the Hamiltonian H_{mf} for $B = 6\,\mathrm{T}$ and $n = 40$ manifold of the field free case.

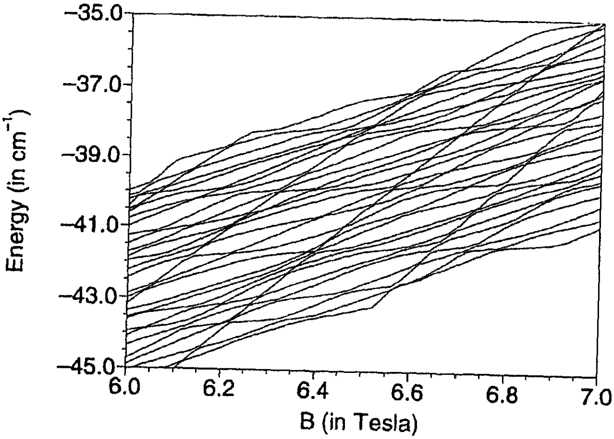


Figure 4. The energy levels belonging to the 410th to 438th eigenstates of a hydrogen atom in a magnetic field associated with the Hamiltonian H_{mf} for the range $B = [6\,\mathrm{T}, 7\,\mathrm{T}]$.

energy value that corresponds to the $n = 40$ manifold of the field free case. Quantal spectra for magnetic field strengths between 6 T to 7 T are shown in figure 4. We have included 29 energy levels between the energy range of $E = -45.0\,\mathrm{cm}^{-1}$ and $-35.0\,\mathrm{cm}^{-1}$. By careful visual inspection, one can notice many avoid crossings arising due to the mixing of adjacent states. These avoided crossings are considered to be the fingerprints of classical chaotic motion [5]. In fact from extensive numerical investigation, it has been found that as the strength of the magnetic field is increased from low to high values, the system gradually makes a transition from ordered to chaotic motion [21]. In fact very accurate results for highly excited states are available now [22, 23]. Chaos in this case is often referred to as soft chaos [24].

2.3 *Hydrogen atom in a parallel electric and magnetic field (Stark-quadratic Zeeman effect)*

After investigating the quadratic Zeeman effect (QZE) in more details, it has now become desirable to find other paradigmatic and experimentally realizable systems which have all the desirable properties of QZE, but contain additional simple perturbations. One such a system can be produced by adding an electric field. The resulting Stark-quadratic Zeeman effect (SQZE) is attracting the interest of theorists [25] and experimentalists [26, 27]. Among the various possible relative orientations of the fields, parallel field alignment alone ensures the survival of the magnetic quantum number (m) as a good quantum number throughout the chaotic regime and hence reduces the phase space dimension of the problem to a manageable size.

The Hamiltonian of a hydrogen atom in a parallel electric and magnetic field configuration is

$$H_{\mathrm{p}} = \frac{p^2}{2} - \frac{1}{r} + \frac{B}{2} L_z + \frac{B^2}{8}(x^2 + y^2) + Fz, \tag{2.10}$$

where F is the strength of the electric field which is applied along the z-axis parallel to the magnetic field. For the parallel field problem an approximate constant of motion Λ_{p} can be found for the weak field limit (in the perturbative regime) [26, 27]

$$\Lambda_{\mathrm{p}} = 4A^2 - 5A_z^2 + 10\beta A_z, \tag{2.11}$$

where $\mathbf{A} = \mathbf{L} \times \mathbf{p} + \mathbf{r}/r$ is the Runge–Lenz vector and $\beta = 12F/5B^2n^4$ measures the relative strength of the electric and magnetic fields. In fact for $F = 0$ (QZE), we get the approximate constant of motion for the magnetic field problem. The total energy within a given n-manifold (neglecting the paramagnetic term) is given as

$$E = -\frac{1}{2n^2} + \frac{1}{16} B^2 n^4 \left[1 + \Lambda_{\mathrm{p}} + \frac{m^2}{n^2}\right]. \tag{2.12}$$

By quantizing Λ_{p} semiclassically, 3 different types of states can be identified. Type I and II states arise from degenerate odd and even librational diamagnetic states (they lie at $\Lambda_{\mathrm{p}} = 25\beta^2$ for $0 \le \beta \le \frac{1}{5}$). Type III states arise from the evenly separated odd and even rotational states of diamagnetic manifold (they lie at $\Lambda_{\mathrm{p}} = 10\beta - 1$ for $\frac{1}{5} \le \beta \le 1$). These theoretical predictions have been checked experimentally for hydrogen and lithium atoms [26, 27].

In order to understand these regimes, in figure 5 we have plotted all the (numerically computed) energy levels of $n = 30$ manifold of the hydrogen atom for F ranging from 0 to 40 V/cm, at magnetic field 2.33 T. Our calculations were performed at unit intervals of F and hence finer details, such as avoided crossings, may be incorrect. All the above mentioned three regions are clearly marked in the figure.

As far as non-perturbative calculations are concerned, Richter *et al* [28] carried out numerical quantal calculations of cases where (i) the magnetic field was dominant, (ii) the electric field was dominant and (iii) both electric and magnetic fields were of comparable magnitude. Surprisingly different n-manifolds overlapped without significant mutual interaction. The avoided crossings of states belonging to adjacent n-manifolds

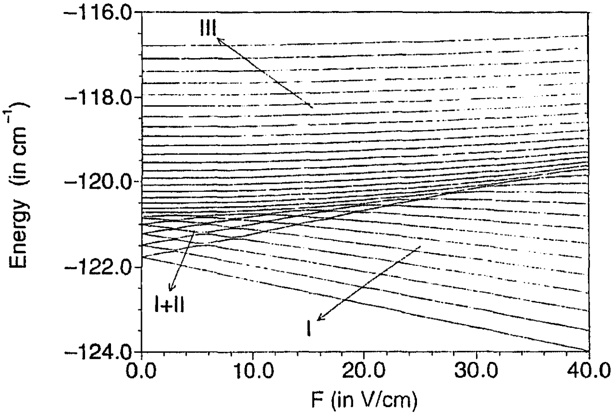


Figure 5. The three distinctive regions of energy levels of $n = 30$ manifold of a hydrogen atom in a parallel electric and magnetic field associated with the Hamiltonian H_p for $B = 2.33\,\mathrm{T}$ and E over the range [0–40] V/cm.

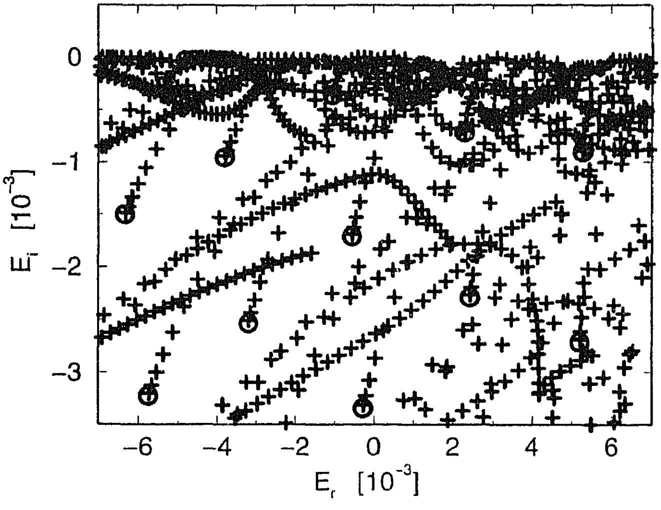


Figure 6. The patternless arrangement of resonances (complex eigenvalues) of a hydrogen atom in a parallel electric and magnetic field associated with the Hamiltonian H_p for $E = 514200\,\mathrm{V/cm}$ and B over the range [0, 1504] T. Open circles belong to $B = 0$ case while + signs belong to $B \neq 0$ cases.

were very small and were beyond graphical resolution. As a consequence of this most of the classical phase space regimes remained regular (of course, all the calculations have been reported around $n = 24$ manifold). We believe that even though the hydrogen atom in parallel fields is a non-separable problem, identifying chaotic regimes seems to be a difficult task [28]; recently the bifurcations of periodic orbits of this system have been reported [29, 30] which may lead to the understanding of manifestation of chaos in this system.

When the energy and field strengths are increased to far beyond the Stark saddle point, one can observe irregularity, particularly in the arrangement of resonances in the complex energy plane [8,9]. In figure 6 (ref. [8,9]) the resonances for a fixed electric field $F = 514200$ V/cm and magnetic fields B varying from 0 T to 1504 T are shown. Resonances for $B = 0$ case are indicated as open circles, and for $B \neq 0$ as plus signs. The overall picture looks quite complicated as the resonances are arranged in a patternless fashion, contrary to the electric field case (figure 2).

2.4 *Hydrogen atom in crossed electric and magnetic fields*

Among the hydrogen atom in static (time independent) external field problems, the case of crossed electric and magnetic fields is considered to be the most difficult (at least for theoretical investigations) [31]. Here the magnetic field is usually applied along the z axis while the static electric field is applied along the x axis. The Hamiltonian can be cast in the form,

$$H_c = \frac{p^2}{2} - \frac{1}{r} + \frac{B}{2}L_z + \frac{B^2}{8}(x^2 + y^2) - Fx. \tag{2.13}$$

The crucial point to note down here is that the z component of angular momentum, namely L_z is not a conserved quantity, so the problem cannot be reduced to 2 dimensions in a straightforward way as in the case of parallel electric and magnetic fields. Moreover, like the Stark effect case, the dynamics is scattering above the (ionization) threshold value, and quasi–bound below it.

In recent experimental investigations, a class of quasi-Landau resonances in the spectra of rubidium Rydberg atoms in crossed electric and magnetic fields have been observed [32]. These set of resonances are associated with a small number of *planar* periodic orbits, so it is believed that even the planar limit, i.e, $z = p_z = 0$ of the above problem may reveal interesting dynamical behaviour [33].

Using simple critical point analysis, it is very easy to show that there is a critical point, independent of the magnitude of the magnetic field B, at $E = -2\sqrt{F}$ (for the planar case) which coincides with the approximate Stark saddle point discussed in § 2. In our investigations we stick ourselves to the planar system.

For convenience we scale the position and momentum variables, $\mathbf{r} \rightarrow B^{-2/3}\mathbf{r}$ and $\mathbf{p} \rightarrow B^{1/3}\mathbf{p}$. The advantages of using such scaling relations are described later in detail in § 4. In terms of scaled variables the above Hamiltonian becomes [34]

$$H_c^s = EB^{-2/3} = \frac{p^2}{2} - \frac{1}{r} + \frac{1}{2}L_z + \frac{1}{8}(x^2 + y^2) - FB^{-4/3}x. \tag{2.14}$$

In figure 7 we have sliced the classical phase space that corresponds to $E_s = -1.0$ using $y = 0$ and $p_y =$ arbitrary plane. Whenever the classical trajectory pierce through this plane we record the x and p_x values and obtain the so called Poincaré surface of section (PSS) for a set of initial conditions. To start with, we choose the value of f well below the critical point, $f = FB^{-4/3} = 0.2$. Each set of initial condition leads to an invariant curve on the surface of section which indicates that the motion is quasi periodic in the actual phase space and hence the resulting dynamics is largely regular.

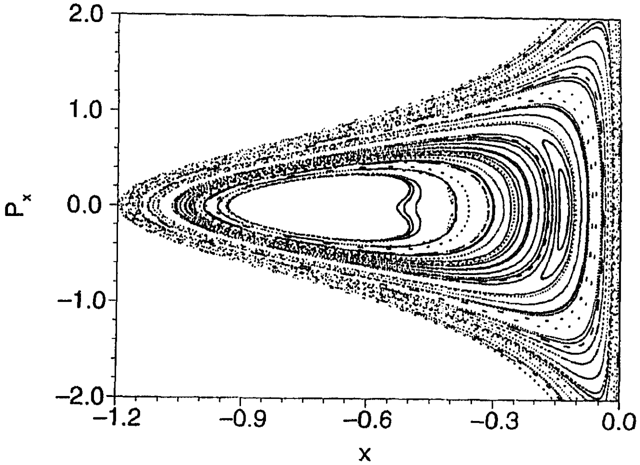


Figure 7. The Poincaré surface of section for various initial conditions of a hydrogen atom in a crossed electric and magnetic field associated with the Hamiltonian H_c^s for the scaled energy $E_s = -1.0$ and scaled electric field $f = 0.2$.

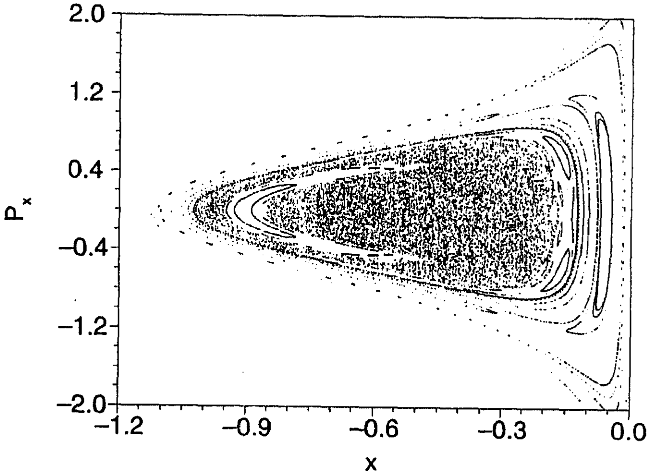


Figure 8. The Poincaré surface of section for various initial conditions of a hydrogen atom in a crossed electric and magnetic field associated with the Hamiltonian H_c^s as in figure 7 but for $f = 0.2499$.

However if we increase the value of f, when f approaches the critical value, we observe chaotic motion. In figure 8 we show a representative PSS plot for $f = 0.2499$ and $E_s = -1.0$. Some initial conditions lead to invariant curves while many others lead to chaotic motion: a sizable fraction of the phase space is occupied by chaotic trajectories.

If we increase f further and cross the critical value, the system enters the scattering regime [33], giving a 'hole', appearing in certain regions of the PSS plot. As f increases, the hole region expands in size. The origin of this hole is attributed to the velocity dependent, Corriolis like force in the Newton's equations of motion (transparent when written down in semi-parabolic coordinates), arising due to the non conservation of L_z in the planar system [33].

Thus, unlike the parallel field problem, as the electric field strength is increased from far below critical point to just below the critical point, the system undergoes a clear transition from regular to chaotic motion, even in the 2 dimensional planar case.

As far as the 3 dimensional case is concerned, if we increase the electric field far above the critical point, resonances (complex eigenvalues) start to appear. If we increase the magnetic field also sufficiently, a large number of resonances begin to appear. In fact, the nearest neighbour spacing distribution of the complex eigenvalues which belong to the chaotic scattering region follow the so-called Ginibre distribution given as [35]

$$P(s) = 2\left[\frac{9\pi}{16}\right]^2 s^3 \exp\left[-\frac{9\pi}{16}s^2\right]. \tag{2.15}$$

At the same time the classical trajectory calculations reveal fractal structures in classical chaotic ionization in these regimes [35].

3. Hydrogen atom in other time independent external fields

So far we have discussed the chaotic dynamics of hydrogen atom in static electric or/and magnetic fields. Now we will consider some other interesting time independent external fields which can be realized physically.

3.1 *Hydrogen atom near a metal surface*

The force between an atom and a nearby metal surface is given by the so-called (instantaneous) van der Waals force. The Hamiltonian of a hydrogen atom near a metal surface can be derived using the method of images as [36]

$$H_{\mathrm{m}} = \frac{p^2}{2} - \frac{1}{r} - \frac{1}{4d} + \frac{1}{\sqrt{x^2 + y^2 + (2d+z)^2}} - \frac{1}{4(d+z)}. \tag{3.1}$$

Here d is the atom-surface distance. When the atom is at a larger distance in comparison to its size ($d^2 \gg r^2$), the above Hamiltonian simplifies to the so called instantaneous van der Waals Hamiltonian H_{ivp} given as

$$H_{\mathrm{ivp}} = \frac{p^2}{2} - \frac{1}{r} - \frac{1}{16d^3}(x^2 + y^2 + 2z^2). \tag{3.2}$$

The above Hamiltonian H_{ivp} is particularly useful near the ground state when the size of the atom is quite small, whereas the Hamiltonian H_{m} is quite useful when dealing with Rydberg states for which the size of the atom is relatively big. In figure 9 we have shown the quantum spectrum of a hydrogen atom over an atom-surface distance range of 100 nm to 300 nm. The calculations were performed for $n = 20$ Rydberg states using the Hamiltonian H_{m}. A number of avoided crossings indicate that there can be chaos at the classical level in this regime. On the other hand, if we use Hamiltonian H_{ivp} and also consider low-lying states, no such avoided crossings appear [37]. Moreover in both the

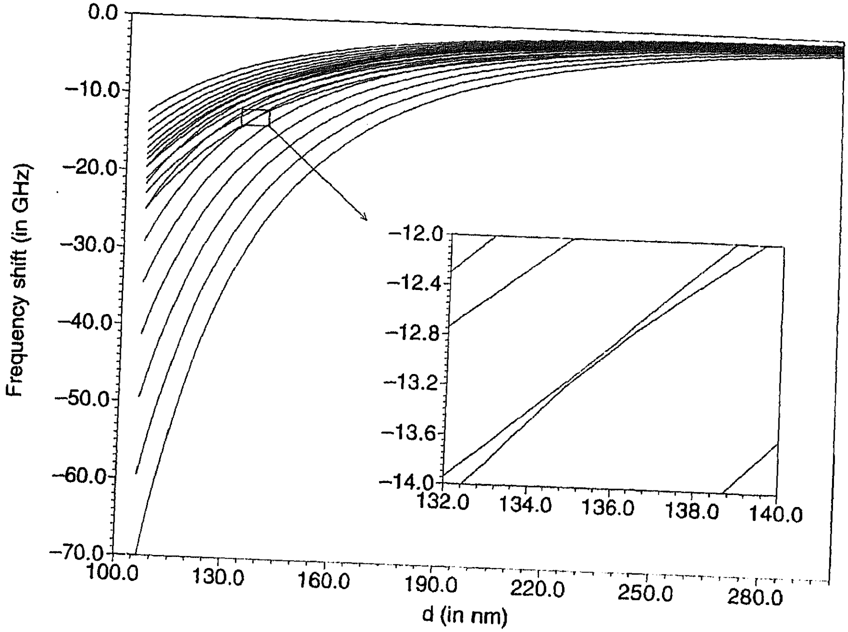


Figure 9. The frequency shifts of the eigenvalues of the $n = 20$ manifold of the Hamiltonian given by (3.1) from the field-free position for an atom-surface separation range of 100–300 nm.

ases the adjacent n-manifolds are well separated and hence do not interpenetrate. Detailed numerical investigations imply that the dynamics is regular near the ground tates and at larger atom-surface distances. If we consider highly excited states and maller atom-surface distances then chaos might appear. However, extensive classical tudies are yet to be carried out in this régime.

.2 *Hydrogen atom in a generalized van der Waals potential*

y generalizing the instantaneous van der Waals Hamiltonian H_{ivp} and the quadratic eeman Hamiltonian H_{mf}, we get the so called generalized van der Waals (GVP) amiltonian given as

$$H_{\mathrm{gvp}} = \frac{p^2}{2} - \frac{1}{r} + \gamma\,(x^2 + y^2 + \beta^2 z^2). \tag{3.3}$$

properly defining γ and β appropriately, we get a variety of interesting physical oblems [38]. Prominent among them are: (i) quadratic Zeeman, (ii) spherical quadratic eman, (iii) instantaneous van der Waals and (iv) the hydrogen atom in parallel plate ve guide problems. Moreover just by reversing the sign of the Coulombic term in the ove Hamiltonian, namely $-1/r$, we obtain the celebrated Paul trap Hamiltonian used the precision atomic spectroscopists for ion confinement [39].

The Hamiltonian H_{gvp} cannot be solved exactly for arbitrary β values. By using indrical and semiparabolic coordinates one can transform the above problem into that

Table 1. Classical and quantal results of hydrogen atom in a generalized van der Waals potential for various β values.

β	Classical phase space	Quantal NNS distribution
0.25	Predominantly chaotic	Wigner distribution
0.5	Completely regular	Poisson distribution
$\sqrt{0.4}$	Both regular and chaotic	Intermediate distribution
1.0	Fully regular	Poisson distribution
1.5	Both regular and chaotic	Intermediate distribution
2.0	Completely regular	Poisson distribution
3.0	Predominantly chaotic	Wigner distribution

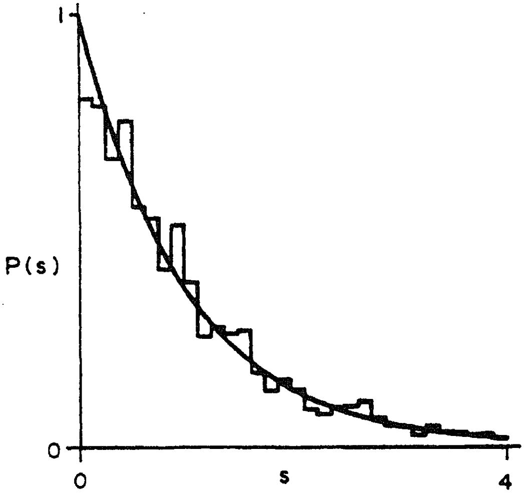


Figure 10. The NNS distribution of a hydrogen atom in a generalized van der Waals potential associated with the Hamiltonian H_{gvp} for $\beta = \frac{1}{2}$.

of a two coupled sixth-power anharmonic oscillators problem. Then by using the Painleve singularity structure analysis and dynamical symmetry approach it is possible to show that the Hamiltonian H_{gvp} is integrable for $\beta = \frac{1}{2}$, 1 and 2 cases only. Then all the recently developed (numerical) techniques to deal with chaotic, nonintegrable Hamiltonian systems can be used to show that there is a choas-order-chaos-order-chaos-order-chaos type of transition. In fact it is possible to perform Einstein–Brillouin–Keller (EBK) type of (semiclassical) quantization and obtain exact analytical expression for the energy level shifts only for the integrable cases namely $\beta = \frac{1}{2}$, 1 and 2. Other near integrable cases can be quantized using numerical methods. All our semiclassical predictions were on par with our classical results [40]. Some of the classical and quantal results have been summarized in table 1.

As already stated, the nearest-neighbour spacing distribution (NNS) (the probability P for finding a separation s of neighbouring energy levels in the spectrum) of a quantal system follow the Poisson distribution i.e, $P(s) = \exp(-s)$ when its classical counterpart exhibits regular behaviour. Quantal systems follow the so-called Wigner distribution (or

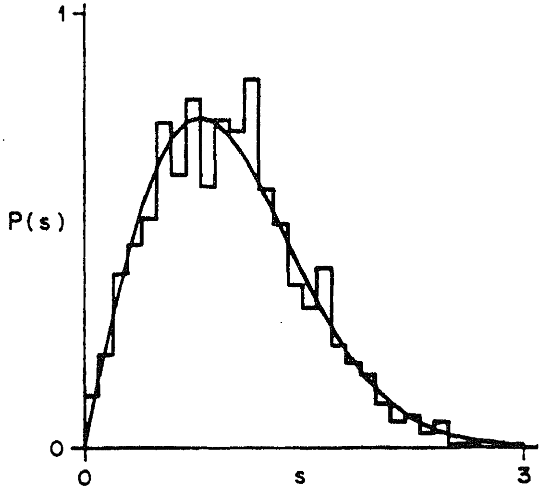


Figure 11. The NNS distribution of a hydrogen atom in a generalized van der Waals potential associated with the Hamiltonian H_{gvp} for $\beta = 3$.

Gaussian Orthogonal Ensemble distribution in our case of time independent system) when the corresponding classical system exhibits chaotic behaviour. The quantum problem associated with the above Hamiltonian H_{gvp} can be analysed by solving the associated Schrödinger equation numerically in terms of the $SO(4,2)$ group generators using a scaled set of normalized Coulombic wavefunctions. We get a large number of converged eigenvalues and they can be used for the level statistics analysis. As the parameter β is varied, we obtained a GOE–Poisson–Brody–Poisson–Brody–Poisson–GOE type of transition corresponding to classical dynamics. In figures (10) and (11) we show the NNS distribution of the generalized van der Waals Hamiltonian for $\beta = \frac{1}{2}$ and 3 which correspond to classically regular and chaotic situations.

The detailed study of generalized van der Waals problem implies that the system makes a chaos-order-chaos-order-chaos-order-chaos type of transition when the β parameter is varied over the range $[0, 3]$. In fact similar conclusions have been reported in the case of Paul trap Hamiltonian problem [39]. Some of the ion trap experiments have verified these results [41].

3.3 *Anisotropic hydrogen atom problem*

So far we have discussed systems in which the nonlinearity was brought in by the interplay between the Coulomb field and the external electric or/and magnetic field(s). Now we consider the motion of a charged particle with an anisotropic mass tensor [42–44]. Let the mass of the electron along x and y directions be m_{p} while that along the z direction be $m_{||}$. Then the mass ratio $\gamma = m_{\mathrm{p}}/m_{||}$ is used to define the so called mass anisotropy as $(1 - \gamma)$. In reality different semiconducting materials have different effective mass (γ) values. For example using cyclotron resonance work it has been found that the γ value for silicon and germanium are respectively 0.2079 and 0.05134 [45].

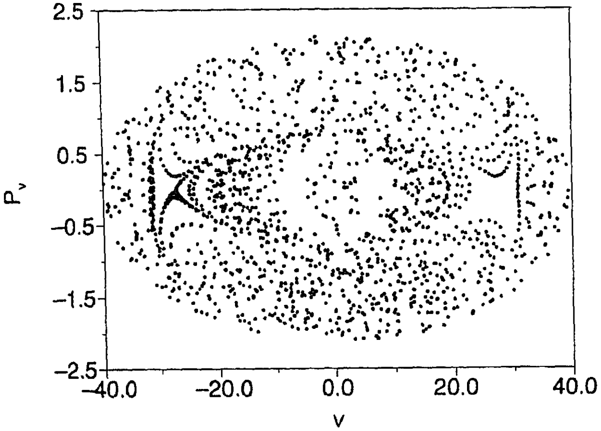


Figure 12. The Poincaré surface of section for a single initial condition of an anisotropic hydrogen atom associated with the Hamiltonian H_a for $\gamma = 0.85$.

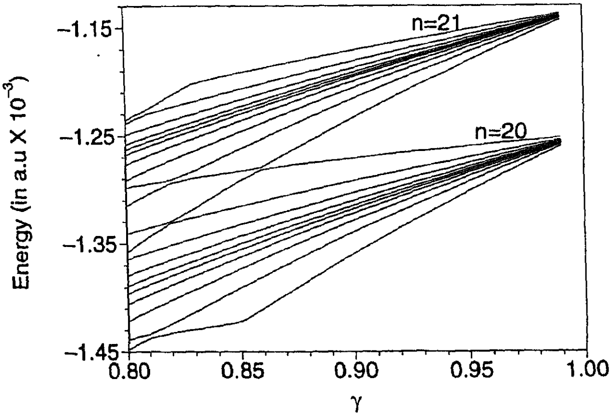


Figure 13. The energy levels belonging to $n = 20$ and 21 manifolds of the anisotropic hydrogen atom associated with the Hamiltonian H_a for γ over the range [0.8, 1.0].

For simplicity we consider the hydrogen atom and introduce the mass anisotropy. The Hamiltonian can be cast in the form

$$H_a = \frac{p_x^2 + p_y^2}{2} + \frac{\gamma}{2} p_z^2 - \frac{1}{r}. \tag{3.4}$$

Here $\gamma = 1$ corresponds to the isotropic situation. We gradually decrease γ from unity and study the dynamics: for $\gamma = 1$ (isotropic case), the system is integrable while for $\gamma < 1$ the system abruptly becomes chaotic. As the accessible phase space in the Cartesian coordinate space is quite awkward for visual inspection, we have used the semi parabolic coordinates which are defined as $\rho = \sqrt{x^2 + y^2} = uv, z = (u^2 - v^2)/2$. In these calculations for the classical trajectory can be carried out in a nice fashion. The Poincaré surface of section plot (for $u = 0$ and $p_u > 0$ plane) in the v versus p_v plane for $\gamma = 0.85$ and an energy value that corresponds to $n = 20$ manifold of the isotropic case is shown in figure 12. Even a single trajectory is able to traverse almost the entire accessible phase space and hence the $v - p_v$ plane is filled with randomly distributed points, implying that the system becomes completely chaotic.

In figure 13 we plot the quantal eigenvalues for $n = 20$ and $n = 21$ manifolds, for γ in the range [0.8, 1.0]. In the isotropic case namely $\gamma = 1$ the adjacent manifolds are well separated. As the anisotropy is gradually introduced, the adjacent n-manifolds interpenetrate, which might be due to the existence of chaotic motion at the corresponding classical system. Unlike the other systems discussed so far, here the transition from regular to chaotic motion is considered to be abrupt [43] and is often referred to as hard chaos [24].

4. Hydrogen atoms in strong oscillating electromagnetic fields

4.1 *Early experiments and theory*

The hydrogen atom interacting with a strong, periodic monochromatic electromagnetic field is an example of another nonlinear system, which in a certain range of parameters may display signatures of classically chaotic behaviour.

Classical and quantal analytical solutions are very well known separately for both, namely the field-free hydrogen atom and the free electron in the electromagnetic field (see e.g. [11]). Since the motion of the electron is governed by the internal Coulombic and external laser field, it is expected that chaotic dynamics may be observed when both internal and external fields have comparable effect on the electron. Therefore in order to create 'strong fields' and introduce conditions beyond the perturbative regime for the ground state or low lying states we will have to apply ultrashort intense pulses of laser light, whereas for high Rydberg levels the required electric fields may be quite small in terms of the absolute value. Nevertheless, the experiments with Rydberg atoms are very difficult due to the presence of 'stray' electric fields which may easily ionize these atoms. Stray (unknown) electric fields originate from electric charges present in the experimental apparatus. At the same time, the experiments with ultra-short pulses are by no means easier—it is extremely difficult to obtain a good repetition rate of parameters of the laser pulse and it is also difficult to measure their peak intensity accurately [46].

Rydberg hydrogen atoms interacting with microwave field have been of primary interest for over 20 years [47]. Microwaves have a dramatic effect on excited hydrogen because the splitting between levels n and $(n+1)$ is approximately $\delta E(n, n+1) \approx 6.6 \times 10^6\, n^{-3}$ GHz, which is in the microwave region for $n \sim 50$–100. Thus microwaves cause resonant transitions between adjacent levels over a wide range of Rydberg states. In a pioneering microwave experiment [47], hydrogen atoms were initially prepared in a Rydberg state with principal quantum number $n_0 \approx 66$ and then were introduced into the interaction region (microwave cavity) with the microwave field of frequency $\nu = 9.9$ GHz. Although for this case about $N_f = 80$ photons were to be absorbed in order to reach the ionization limit from the initial state in the perturbative alkali atoms [50].

Typically in the ionization experiments with Rydberg atoms one measures the electric field ionization threshold, F10%, which is an amplitude of the electric field for which 10% ionization probability occurs. It has to be stressed that both true ionization and the excitation to above a certain n-cutoff, n_c ($n_c > n_0$ is determined experimentally) contribute to the experimental value of the ionization field threshold for the atom prepared initially in the state n_0. This is due to the presence of the stray electric fields, which would necessarily cause the ionization. Therefore experimentally one is not able to tell the difference between the true ionization and the excitation higher than n_c [24]. In order to explain experimental results classical Monte–Carlo type theories were applied to the problem [51]. The set of classical initial conditions were chosen so as to mimic the quantum wavefunction probability distribution and the ionization probability was determined as a ratio of the number of ionizing initial conditions to the number of all chosen initial conditions. These simulations have shown that the ionization threshold is related to the onset of chaotic motion of trajectories, which in the presence of critical threshold field become unbounded in the phase-space.

4.2 *Classical model*

Let us write the Hamiltonian of the hydrogen atom in a laser field of an arbitrary polarization (in the dipole gauge approximation)

$$H = \frac{p_x^2 + p_y^2 + p_z^2}{2} - \frac{1}{r} - Ff(t)[x\cos(\omega t + \varphi) + y\tan(\chi)\sin(\omega t + \varphi)], \quad (4.1)$$

where $r = \sqrt{x^2 + y^2 + z^2}$ and F, ω, φ are respectively the electric field amplitude, field frequency and initial phase of the field. We do not include influence of the magnetic field because the associated Lorentz force is much smaller than the Coulombic force for non-relativistic field intensities (less than $10^{18}\,\mathrm{W/cm^2}$—[52]) as the Rydberg electron velocity $v \ll c$. Through the function $f(t)$, the pulse envelope, the turn-on and off of the pulse of length T over finite time, τ may be included,

$$f(t) = \begin{cases} \sin^2(\pi t/2\tau), & 0 < t < \tau \\ 1, & \tau < t < T - \tau \\ \cos^2[\pi(t - (T - \tau))/2\tau], & T - \tau < t < T \\ 0, & \text{elsewhere.} \end{cases} \quad (4.2)$$

The angle $\chi \in [0, \pi/4]$ parametrizes the degree of ellipticity of the field polarization: $\chi = 0$ for linear polarization (LP) and $\chi = \pi/4$ for circular polarization (CP) in particular. Notice that in the case of $\omega \to 0$ we have the well known Stark effect described earlier in § 2.

$$\begin{aligned}
\mathbf{r} &\to \lambda^{-1}\mathbf{r},\\
\mathbf{p} &\to \lambda^{1/2}\mathbf{p},\\
t &\to \lambda^{-3/2}t\\
\omega &\to \lambda^{3/2}\omega,\\
F &\to \lambda^{2}F,
\end{aligned}\tag{4.3}$$

for any real parameter λ. Considering the hydrogen atom with energy $E_n = -1/(2n^2)$ in the electric field of amplitude F and frequency ω, suggests setting $\lambda = n^2$, and consequently $\omega_0 = n^3\omega$ and $F_0 = n^4F$. Thus it is possible to study only the classical dynamics of the atom with initial energy $E = -1/2$ irrespective of the initial state n, in terms of the scaled variables: E, ω_0 and F_0. Non-scaled values of the parameters may be reconstructed from the scaled variables for any arbitrary n_0. The scaled frequency ω_0 can be interpreted as non-scaled frequency expressed in units of the Kepler frequency $\omega_K = (-2E_n)^{-3/2}$ of the electron on its initial orbit.

However, the commutation rule is not preserved under the scaling transformation eq. (4.3):

$$[x, p_x] = i\hbar \xrightarrow{\text{scaling}} [x, p_x] = i\hbar \cdot \lambda^{-1/2} = i\hbar/n. \tag{4.4}$$

Nevertheless, as the semiclassical limit of $n \to \infty$ (effective Planck constant $\hbar/n \to 0$) is approached, the commutation rule $[x, p_x] \to 0$ becomes less affected by the scaling transformation. Again, we see the reason why Rydberg levels are of great interest. The investigation of their properties helps us to understand the relation between the classical ($\hbar = 0$) and the quantum mechanics ($\hbar = 1$ in atomic units) and the way of reaching the semiclassical limit ($\hbar \to 0$) in particular. For the low frequency limit $\omega_0 \leq 0.2$ the classical scaling is visible quite clearly in experimental (quantal) results for the ionization threshold field when $24 \leq n \leq 56$ [50]. It supports the validity of the classical approach to the problem, which in many respects is much easier than the quantum case.

As in the case of static magnetic field, we transform the Hamiltonian equation (4.1) into the problem of two coupled oscillators using semiparabolic coordinates [21] (2D) or using Kustaanheimo–Stiefel (KS) transformation to the problem of four coupled oscillators with a common frequency. The (KS) transformation [38, 53, 54] exploits the $SO(4)$ symmetry of the hydrogen atom. Then a new Hamiltonian is used to write a set of equations of motion, which form a set of coupled differential equations. The transformation to the oscillator problems is advantageous, because the equations of motion are free from the Coulomb singularity and hence can be integrated (solved) numerically to a very high accuracy. Initial conditions are chosen according to the microcanonical ensemble [55]. This ensemble represents a subspace of constant energy (corresponding to initially field-free atom) of the phase-space which is formed by generalized coordinates and canonically conjugated momenta. Then Monte-Carlo type of calculations are made and the ionization probability, P, is obtained as a ratio of the number of initial conditions which lead to the ionization to the total number of initial conditions.

As it was mentioned above, classically, the ionization occurs due to the break up of the Kolmogorov–Arnold–Moser (KAM) tori when the microwave amplitude is large enough. Hence the ionization threshold can be associated with the onset of classical chaos. Therefore, in the presence of the external periodic force we can observe escape of the electron to the continuum from a certain phase-space region defined by the initial conditions. For Hamiltonian systems, the escape may be significantly slowed down due to the presence of the remnants of KAM tori (Cantori) resulting in a power-law (algebraic) decay $S \sim t^{-z}$ from the region containing KAM stability islands, rather than exponential decay $S \sim \exp(-\gamma t)$, in the limit of large time t. This slow-down may be greatly enhanced in the quantum description because of the coarse graining of the phase-space [56, 57] due to the finite value of $\hbar$. To get more insight into the exponential and power-law decay, let us consider a mathematical model of the ionization [58]. First, let us consider the canonical Cantor set (the "middle-third" set). We take interval of the unit length. In the first step we reject $g = \frac{1}{3}$ fraction of points from the middle of the interval, thus creating 2 smaller intervals of length $\frac{1}{3}$ each. In following steps, repeated *ad infinitum*, we just apply the above procedure to each newly created interval in the preceding step, obtaining finally the Cantor set, which is a well known fractal object. The survival probability (of not being deleted), s, of the point from the original unitary interval after $N+1$ steps is equal to the Lebesgue measure of the remaining set of points:

$$s(N+1) = (1-g)s(N) \tag{4.5}$$

and consequently, it can be easily verified that the resulting survival probability is an exponential function of the step N (which is a discretized equivalent of time t):

$$s(N) = \exp(-N\gamma), \tag{4.6}$$

where $\gamma = -\ln|1-g|$ is positive (if $g < 1$). Modifying the Cantor set construction, and assuming that the fraction deleted, g_N, is not constant, but decreases from one step to another eg.: $g_N = g/N$, we can write the survival probability

$$s(N+1) = (1-g_N)s(N). \tag{4.7}$$

For large N one can check that a power-law dependence

$$s(N) \approx N^{-g}, \tag{4.8}$$

fulfills the iterative equation (4.7)

Indeed, such a power-law behaviour of the survival probability s has been observed in physical situations approached by a one-dimensional (1D) model of hydrogen interacting with the LP field, included as a periodic train of very short perturbations (the so called kicked hydrogen model) [58]. This behaviour has been supported by a 1D model of hydrogen with a monochromatic type of excitation [59]. Classical simulations indicate that both in the LP [60] and the CP field [61, 62] problem, this sort of long time behaviour could be observed. Quite recently, both experimental and quantum (numerical) evidence has been provided for the existence of the power-law (algebraic) decay over several orders of magnitude [62].

In the quantum approach to the problem we have to solve the time-dependent Schrödinger equation

$$H|\psi(t)\rangle = i\hbar\frac{\partial|\psi(t)\rangle}{\partial t}. \tag{4.9}$$

One may directly integrate this time dependent equation. It is however only possible to do so for pulses (interaction times) no longer than 30–50 field cycles [63, 64], because of the increasing importance of interaction of the wave function with the edges (the so called absorbing boundary conditions) of the integration grid. Alternatively we may take an advantage of periodic time dependence in the Hamiltonian equation (4.1). We represent the solution of the Schrödinger equation in the Floquet–Fourier expansion as [65]

$$|\psi(t)\rangle = \sum_j a_j \,\mathrm{e}^{-iE_jt}|\psi_j(t)\rangle. \tag{4.10}$$

We may express each component $|\psi_j(t)\rangle$ in terms of the multiphoton expansion

$$|\psi_j(t)\rangle = \sum_K \mathrm{e}^{-iK\omega t}|\psi_j^K\rangle. \tag{4.11}$$

We substitute this to the Schrödinger equation and the resulting Floquet Hamiltonian H_F [65], which is time-independent, describes "dressed-atom" model. Thus instead of discrete states we have resonances (complex energies). The time evolution operator U for Floquet Hamiltonian H_F acting in the 'photon-space' is written in the following form [65]

$$U = \exp(-iH_\mathrm{F}t/\hbar) = \frac{1}{2\pi i}\int_C \mathrm{d}z\frac{\exp(-izt/\hbar)}{z-H_\mathrm{F}}. \tag{4.12}$$

One can see that the time evolution is dominated by poles of the operator $(z-H_\mathrm{F})^{-1}$ near the real axis in the complex energy plane. These complex poles (resonances) may be found from the analytical continuation of the Floquet Hamiltonian $H_\mathrm{F}(\Theta)$, which is obtained by using the dilatation transformation [65, 66]

$$\begin{aligned} r &\rightarrow r\,\mathrm{e}^{i\Theta}\\ p &\rightarrow p\,\mathrm{e}^{-i\Theta}. \end{aligned} \tag{4.13}$$

The eigenenergies (quasi-energies) E_j of the H_F are found usually in the harmonic oscillator basis representation [66] and if $\tau \ll T$ (a flat top pulse with narrow edges of turn-on and off time—a rectangular pulse shape approximation (RPSA)) the ionization probability can be expressed as [62]

$$P_\mathrm{RPSA} = 1 - \sum_j |c_j|^2 \mathrm{e}^{-\Gamma_j T}, \tag{4.14}$$

where T is the interaction time (length of the pulse). The case with a smooth turn–on could be dealt with on the grounds of the single state approximation (SSA) [67]. Γ_j [$\Gamma_j = -2\Im(E_j)$] is a field-induced width of the jth Floquet eigenstate, which accounts for there is only one dominant overlap c_j ($c_j \approx 0.9$) of a Floquet eigenstate on the initial state Φ_0 (describing the initial Rydberg state), in the sum eq. (4.14), then we may well apply the SSA in which only this Floquet state of width Γ may be included to obtain the total ionization probability

$$P_{\mathrm{SSA}} = 1 - \exp\left\{ - \int_0^T \Gamma[F_0 f(t)] \mathrm{d}t \right\}. \tag{4.15}$$

In the following subsections we will consider special cases of the hydrogen atom interacting with the LP and CP fields. Then we will show some typical results for the CP case using the above mentioned classical and quantum methods.

4.4 *LP microwave field problem*

In the LP case, since the projection of the angular momentum operator onto the polarization axis is conserved, we may consider a general case which is effectively two-dimensional (2D). The PSS portrait of the classical phase-space is subdivided into regions from which a classical trajectory cannot escape due to the presence of invariant KAM tori [2]. If the perturbation parameter (electric field) exceeds a certain critical value, F_{cr}, KAM tori break up, resonances begin to overlap [68] and trajectories may enter regions previously non-accessible to them. This property of 2D phase-space (and 1D as well) may be of importance if one is interested in investigating global diffusion as a signature of the global chaotic behaviour.

Using theoretical and experimental methods one can find the ionization field threshold in a vast region of scaled frequencies. The experiments for hydrogen covered the scaled frequency ω_0 region from 0.021 to 2.8. It has been achieved by choosing initial states n_0 between 24 and 90 for fixed frequencies from the interval of 7.58–36.02 GHz [24]. Thus for $\omega_0 \approx 1$ atoms were subject to pulses of length 300–400 external field oscillations. Results of the experiments allowed to determine regimes of dynamical behaviour of the system and discuss mechanism of the ionization. We refer the reader to ref. [24] for detailed discussions and here we just sketch a rough description of main frequency regions. For very low frequencies (much below the Kepler frequency) we have a situation similar to the static homogeneous field, and classically ionization occurs as an escape above the potential barrier. The motion is quasi-regular and quantal corrections due to the tunneling may be included through semiclassical methods [69]. In the high frequency regime (much greater than the Kepler frequency $\omega_0 > 1$) it is possible to obtain an effective classical potential with respect to the fast field oscillations [54, 70]. In the high frequency regime quantum results indicate much higher field threshold for the ionization than the classical simulation, which is referred to as the so-called quantum photonic localization [71]. Classical aspects of quantum localization in microwave ionization has been discussed recently [72] and it has been found that part of the discrepancies with quantum predictions may be removed if the Coulomb singularity is ‘softened’ [72]. In the intermediate region of frequencies, excitation occurs due to the diffusive gain of energy which finally leads to the ionization [24, 71]. This is the main ionization mechanism in this region, but some local maxima on the field threshold dependence on the frequency can occur due to the stabilizing influence of classical resonances [24].

In the theory of hydrogen atoms interacting with the LP field the classical Kepler model and its corresponding quantum counterpart are of paramount importance [71,73]. The derivation of the Kepler map is based on the observation that if the field frequency is much higher than the Kepler orbital frequency, the strongest interaction between the classical electron and the oscillating field occurs when the electron is close to the nucleus. Far away from the nucleus the electron is a quasi-free particle and negligible work is done on the electron. Because the interaction between nucleus and the electron in the field occurs only during the short time, when the electron is close to the nucleus, the effect of perturbation could be approximated by an instantaneous impulse (or a 'kick') which appears once every orbital period of motion. Therefore we may iteratively write an averaged over one motion period change of the electron energy and its phase as

$$\begin{aligned} N_{i+1} &= N_i + k\sin(\varphi_i) \\ \varphi_{i+1} &= \varphi_i + 2\pi\omega/(-2\omega N_{i+1})^{3/2}, \end{aligned} \tag{4.16}$$

where $N = -1/(2n^2\omega)$ is the electron energy in units of the photon energy (frequency). N increases or decreases depending on the phase, φ, which in turn depends on the period of the orbit corresponding to a certain n value, $T = 2\pi n^3 = 2\pi/(-2\omega N_{i+1})^{3/2}$. The parameter $k = 0.822\pi F/\omega^{5/3}$ measures the strength of the 'kick' for each orbit. The long time evolution (measured in Kepler periods) of this map could be easily found for field frequencies $\omega_0 > 1$, however, the applicability of this method may be restricted for the situation where the significant ionization occurs.

The 1D Kepler map model proved to be extremely successful in explaining the full 3D behaviour in experimental situation where we have a very 'elongated' ($l \ll n_0$) high Rydberg states in the LP field. The results of the quantum model and comparison with the classical predictions for the high-frequency region ($\omega_0 \gg 1$) showed the discrepancies in predictions of the ionization threshold which then led to the formulation of the photonic localization theory [71].

4.5 *CP microwave field problem*

In the CP case, unlike the LP situation, the magnetic quantum number (projection of the angular momentum onto the quantization axis) is not conserved and there is no other constant of motion except for the weak field limit [74,75]. Thus the CP case needs in general a full 3D treatment which is a huge task for quantum calculations. In the 3D classical model, for an arbitrary small perturbation it is possible to observe the Arnold diffusion [2]. However, it is possible to consider a simpler 2D system, for which KAM tori still divide phase-space into secluded regions. The application of the 2D model could be especially justified if the initial electron orbit plane coincides with the field polarization plane [76].

A transformation of the Hamiltonian eq (4.1) (written for $\chi = \pi/4$) to the rotating frame with the field [76], which is quantum mechanically equivalent to making a unitary transformation $U_{\text{rot}} = \exp(i\omega_0 L_z t)$ [67], yields a new Hamiltonian

$$H_{\text{rot}} = \frac{p_x^2 + p_y^2 + p_z^2}{2} - \frac{1}{r} - \omega_0 L_z - F_0 f(t)x. \tag{4.17}$$

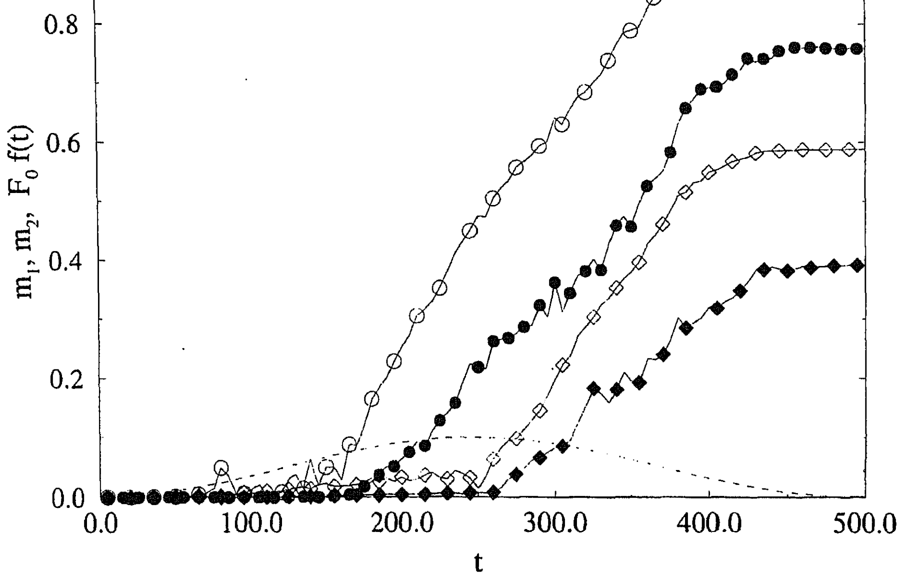


Figure 14. Statistical moments m_1 (open symbols), m_2 (filled symbols) as a function of time. The pulse has length of $T = 500$ field cycles, turn-on and turn-off is made over $\tau = 250$ field cycles. The amplitude envelop of the pulse $f(t)$ is shown with a dashed curve. Circles show data for $\omega = 1.9$ and $F_0 = 0.07$ and diamonds for $\omega = 2.2$ and $F_0 = 0.13$. The field values are taken so as to get similar ionization probability, P, after the pulse ends ($P \approx 0.04$).

Apart from the missing diamagnetic term, this Hamiltonian resembles the Hamiltonian equation (2.13) for the crossed electric and magnetic static field problem, and is time-independent if the change of the pulse envelope is negligible.

Experiments with the alkali atoms [50] have shown that the intensity for the CP case needs to be higher than for the LP case in order to obtain the same ionization probability. Moreover, for hydrogen in a Rydberg state the ionization threshold, F_{th}, scales with the principal quantum number n, and for the LP case the threshold scales according to $F_{\text{th}} = 1/(9n^4)$, whereas for the CP field it scales as $F_{\text{th}} = 1/(6n^4)$ [49]. The latter experimental finding has been supported by theoretical calculations [75]. Note that the scaling of the field amplitude is typical for the classical mechanics (see eq. (4.3)). Intensive studies of the CP case have been recently conducted only for a few years now, despite the fact that early theoretical works [77] started to emerge before the experimental data were even available for the CP case. The ionization threshold in the classical model displays an asymmetry for $\omega_0 = 0$ [76] (a change in the sign of ω is equivalent to the $m \to -m$ change), which is supported by results of the quantum model [67]. Subsequent studies provided an extension of the classical and quantum results to the high frequency region [54, 78]. In the CP problem there were also investigations on the applicability of the Kepler map [79] and the classical resonance overlap criterion [80] to the ionization

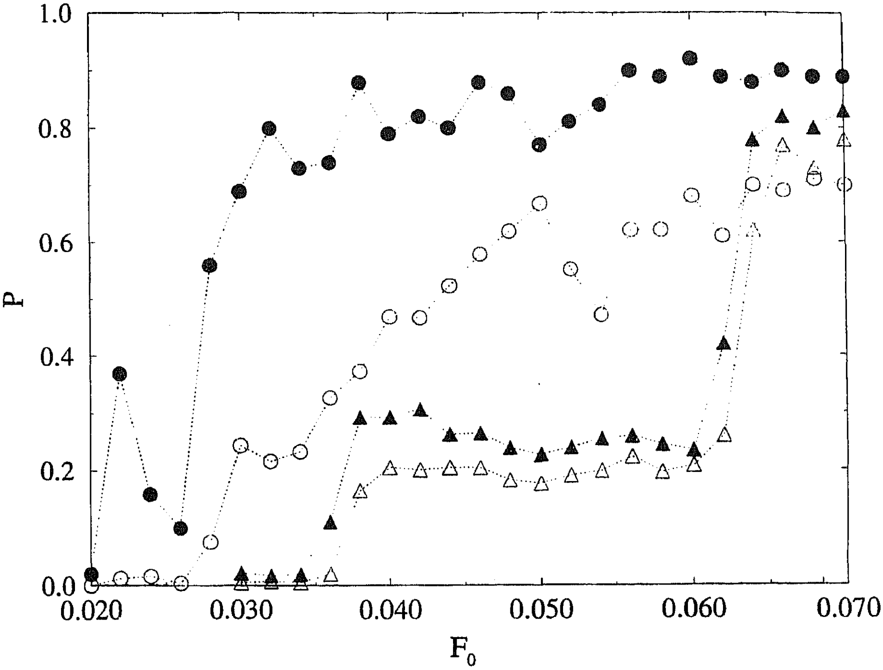


Figure 15. Ionization probability P as a function of maximal amplitude F_0 for long interaction times in the 2D model; $T = 500$ – open symbols and $T = 1500$ Kepler cycles (cycles of electron oscillations on its orbit) – filled symbols; in each case turn-on time $\tau = 50$. Triangles show data for $\omega_0 = 1$, and squares for $\omega_0 = 1.2$.

phenomena. The influence of the pulse shape and initial state preparation on the ionization mechanism in classical [54,61,81,82] and quantum models [78] were also discussed. Recently some stable packets were observed in 2D and 3D quantum models for the CP [83–85], which demonstrate stabilising influence of the classical resonance island. However, it is worth to mention that such non-dispersive packets have been studied theoretically in the LP problem earlier [13].

4.6 *Classical versus quantum model*

We will discuss here some results for the 2D hydrogen atom interacting with the CP field in the intermediate region of frequencies ($\omega_0 \sim 1$), where classical diffusion is the main ionization mechanism. We will consider here only initial conditions which correspond to hydrogen circular states (with the maximum angular quantum number l). The 2D model well approximates full 3D behaviour in particular when initially $|m| = l$ (the classical electron orbit coincides with the CP field polarization plane) [76]. For the sake of demonstration of diffusive behaviour we may consider statistical properties of the ensemble of initial conditions. Let us define statistical moments [54, 86]

$$\begin{aligned} m_1(t) &= \langle n(t) - n_0 \rangle, \\ m_2(t) &= \langle (n(t) - \langle n(t) \rangle)^2 \rangle / n_0^2. \end{aligned} \tag{4.18}$$

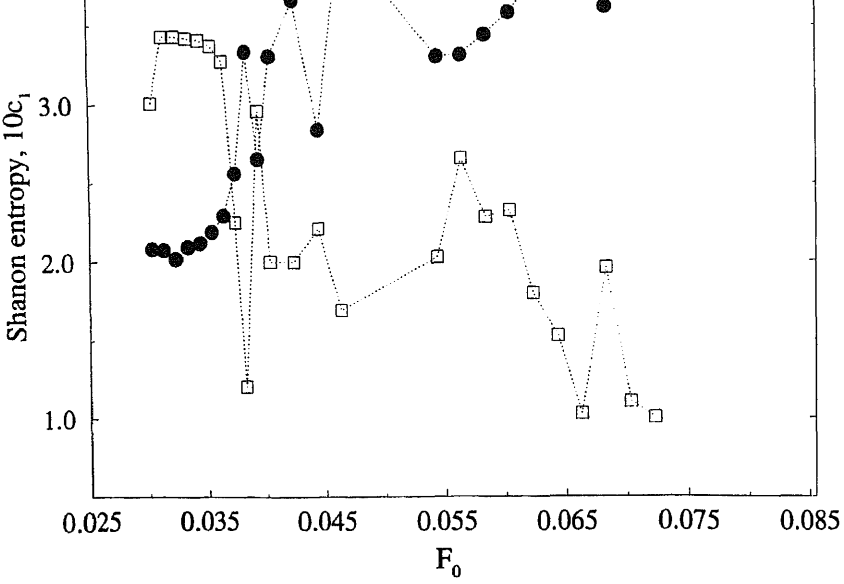


Figure 16. Shanon entropy and maximal overlap onto the initial state coefficient, c_1 as a function of the field amplitude F_0 for the fixed $\omega_0 = 1$. Initial circular state of the field-free hydrogen atom $n_0 = 25$ (field free energy E of the 2D field-free hydrogen defined as $E = -(1/(2(n_0 - 1/2)^2))$ – see e.g. [67]). Note that a jump in value of the Shanon entropy $S = -\sum_j^{j_{\max}} c_j \ln(c_j)$, where $j_{\max}$ number of overlap matrix elements c_j in descending order yielding 99.5% of the total population — $\sum_j^{j_{\max}} c_j = 0.995$.

Here n corresponds to the principal quantum number and is calculated from the energy of the electron $E_n(t) = -1/[2n^2(t)]$. We assume that initially, $n(t=0) = \delta(n_0 - 1)$. In figure 14 we have shown the statistical moments m_1 (open symbols), m_2 (filled symbols) as a function of time. The pulse has length of $T = 500$ field cycles, turn-on and turn-off is made over $\tau = 250$ field cycles. The amplitude envelope of the pulse $f(t)$ is shown with a dashed curve. Circles show data for $\omega = 1.9$ and $F_0 = 0.07$ and diamonds for $\omega = 2.2$ and $F_0 = 0.13$. The field values are taken so as to get similar ionization probability, P, after the pulse ends ($P \approx 0.04$). Note that during the interaction with the CP field the average excitation m_1, as well as its standard deviation (related to square root of m_2) grow.

Figure 15 shows the ionization probability versus the field amplitude for frequency $\omega_0 = 1$ and $\omega_0 = 1.2$ and two interaction times $T = 500$ and $T = 1500$ field cycles. For the frequency $\omega_0 = 1$ the curves for both the time scales exhibit a plateau for a range of values of the field, which is an indication of a stabilizing effect of the strong 1:1 resonance. Irrespective of the interaction time we observe a non-monotonic dependence of the ionization probability on the field amplitude, demonstrating the trapping influence of the classical resonance, and indicates the possibility of non-hyperbolic scattering or power-law dependence of the survival probability on the interaction time [62]. Hence, the

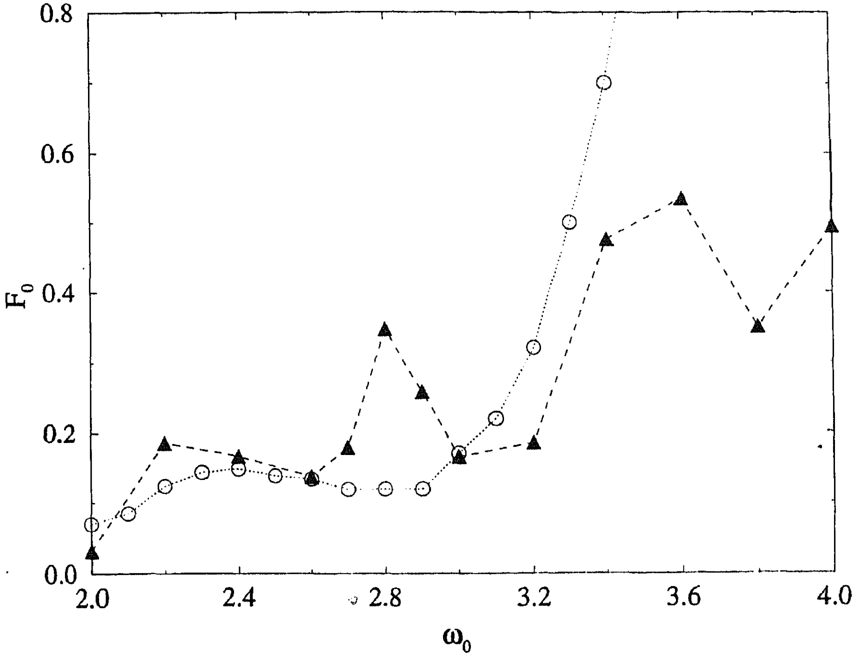


Figure 17. The 10% classical and quantum amplitude threshold in the 2D model. The length of the pulse $T = 500$, turn-on (off) time $\tau = 25$. Quantum results obtained for initial circular state $n_0 = 25$ in the single state approximation are denoted with triangles; clasical data shown with open dots.

ionization is slowed down due to stabilizing structures of phase-space. Moreover, there is numerical evidence for wave packets tracing the classical dynamics *without dispersion,* which are localized in a stable 1:1 classical resonance island [83–85]. For a slightly greater frequency ($\omega_0 = 1.2$) we observe much efficient ionization for the longer interaction time, which again supports the diffusive ionization mechanism.

To get more insight in the corresponding quantum model behaviour, we define the Shanon entropy of Floquet levels, which corresponds to the initial population distribution over a range of Floquet states, obtained for a given value of the field amplitude F_0 and frequency ω_0 as

$$\mathcal{S} = -\sum_{j}^{j_{\max}} c_j \ln(c_j), \tag{4.19}$$

where $j_{\max}$ is the number of overlap matrix elements c_j (calculated as an overlap of the j-th Floquet eigenstate onto the initial state Φ_0) in descending order, which yield 99.5% of the total population — $\sum_j^{j_{\max}} c_j = 0.995$. In figure 16 we depict the Shanon entropy of the overlap matrices for $\omega_0 = 1$ (the 2D model, $n_0 = 25$) in the RPSA approximation equation (4.14). Note a sudden jump in the Shanon entropy around the frequency value for which we observe a growth of the classical ionization probability (see figure 15). Dotted line with squares show value of the maximal, c_1, overlap coefficient (in the figure the values of c_1 are multiplied by 10 in order to use the same scale as for the Shanon entropy). It indicates that for $F_0 > 0.04$ more Floquet states are needed in the expansion equation (4.14) and the initial state is spread over more Floquet states for these values of the field amplitude.

In figure 17 classical and quantum ionization field thresholds are compared for a wider range of scaled frequencies. Note that for the higher frequency $\omega_0 > 3.2$ the classical curve suddenly grows to the high values, which indicates stabilization against the ionization, whereas the quantal results yield a much lower field for the 10% ionization probability.

The model of diffusive excitation leading to ionization has been used in the interaction with linearly and circulary polarized field. Apart from the critical value of the field, $F_{\rm cr}$ (numerically approximated by the $F10\%$ value), it is interesting to study a diffusion time or more strictly speaking an escape time from a certain phase-space region to the infinity [87]. Investigating the long time behaviour may provide us with the information on properties of the classical phase-space. It is therefore arguable, that a 10 % threshold time (length of the interaction/pulse), T10%, is more informative than the field threshold F10% [78]. In particular, the latter may be sometimes ambiguous due to the non-monotonic behaviour ('sub-thresholds') [88] of ionization probability as a function of the field amplitude.

5. Conclusions

In this review we have discussed classical and quantum signatures of the chaotic behaviour of the hydrogen atom interacting with various kinds of physically accessible static and time-dependent external fields. In the case of time independent external fields, we explored bounded classical phase space regions using phase space techniques like trajectory plot and PSS. In the corresponding quantal case, we probed the energy-parameter dependence. For certain cases, we discussed the structure of quantal resonances in the complex energy plane.

As far as the time dependent external field problems are concerned, we have stressed validity of the classical approach for Rydberg initial states. This is supported by the scaling properties of quantum and experimental thresholds with the principal quantum number for low frequency regime, as well as a qualitative agreement between classical and quantum results in the diffusive frequency regime. In the latter regime the onset of classical chaos well approximates the ionization field threshold. We gave an example of the quantum signature of classical chaos in this regime of frequencies and initial conditions, as a sudden growth of the Shanon entropy (defined over the distribution of the initial population over Floquet states). This indicates a complicated level dynamics and abundance of level anticrossings [89, 90], which give rise to redistribution of the initial state over a wide range of Floquet states. This means that the application of the single state approximation for the calculation of the ionization probability becomes more restricted than the rectangular pulse shape approximation.

In table 2, we have summarized both classical and quantal features of hydrogen atoms interacting with external fields. From this table it is evident that the investigation of even a rather simple system, like the hydrogen atom interacting with external fields leads to a variety of interesting and complicated classical and quantal results.

External field	Classical results	Quantum results	Some remarks
Static electric field	Regular motion for any field strength below ionization threshold	Resonances follow regular pattern for field strengths above ionization threshold	Separable and integrable, L_z conserved
Static magnetic field	A smooth transition from regular to chaotic motion if the field strength is varied gradually	Energy parameter space (here B) shows many avoided crossings when the corresponding classical counterpart is chaotic	Non-integrable, L_z conserved; at low fields l-mixing occurs; in the high field limit n-mixing occurs
Parallel electric and magnetic field	Depending upon the electric and magnetic field strengths, regular or/and chaotic motion is expected	Quantal resonances arrange themselves in a patternless manner for strong electric and magnetic fields	Non-integrable, L_z conserved, identification of chaotic regime is non-trivial
Crossed electric and magnetic field	Regular for scaled field strength values well below the critical point and chaotic for very close to the critical point (planar case)	The NNS distribution of quantal resonances follow the so called Ginibre distribution when its classical counterpart is chaotic	Non-integrable, L_z not conserved, Corriolis like force term appears in the Newton's equation of motion; 3D problem
Metal surface	Regular motion for large atom-surface distances; for smaller atom-surface distances chaos is expected	For larger atom-surface distances no avoided crossings in the energy-parameter (here d) space but they appear for smaller atom-surface distances	Non-integrable, L_z conserved, interaction is not strong enough to induce n-mixing of states
Generalized van der Waals field	Chaos-order-chaos type of transitions occur	The NNS distribution shows Wigner–Poisson–Wigner type of statistics	Contains many integrable and non-integrable cases, L_z is conserved
Mass anisotropy	System makes an abrupt transition from regular to chaotic motion when the anisotropy is introduced	Adjacent n-manifolds start to interpenetrate even before l-mixing occurs	Non-integrable, L_z is conserved, Corriolis like term appears in the Newton's equation of motion
Linearly polarized microwave field	Kepler map and 1D models useful, low and intermediate frequency regions results in good agreement with quantum and experiment results	In the high frequency regime indications of quantum photonic localization	L_z is conserved – effectively 2D problem
Circularly polarized microwave field	Antisymmetry around $\omega_0 = 0$ in ionization thresholds; diffusive behaviour for $\omega_0 \sim 1$; stabilizing influence of the 1:1 resonance; stabilization of the atom initially in a circular state for $\omega_0 > 3$	Growth of the Shanon entropy of Floquet levels for the field value where classical onset of diffusive ionization, stable wave packets following classical dynamics	L_z not conserved, 2D model gives good estimates for the ionization thresholds especially when $\|m\| = l$; in rotating frame similar to (static) crossed electric and magnetic field problem

Acknowledgements

The authors would like to thank Profs M Lakshmanan, K T Taylor and J Zakrzewski for their valuable suggestions and comments. Financial support through the award of a post doctoral fellowship by the Queen's University of Belfast (KG) as well as the EPSRC grant, the KBN grant no. 2P03B 03810 and the Stefan Batory Foundation (RG) are gratefully acknowledged.

References

[1] M Lakshmanan (ed.), *Solitons—Introduction and Applications* (Springer-Verlag, Berlin, 1988)

[2] L E Reichl, *The transition to chaos in conservative classical systems: Quantum manifestations* (Springer-Verlag, Berlin, 1992)
A J Lichtenberg and M A Liberman, *Regular and Stochastic Motion* (Springer-Verlag, New York, 1983)

[3] D Delande, in: *Les Houches Session LII, Chaos and Quantum Physics 1989* edited by M-J Giannoni, A Voros and J Zinn–Justin (North Holland, Amsterdam, 1991)

[4] F Haake, *Quantum signatures of chaos* (Springer, Berlin, 1991)
M C Gutzwiller, *Chaos in classical and quantum mechanics* (Springer-Verlag, New York, 1990)

[5] R Ramaswamy and R A Marcus, *J. Chem. Phys.* **74**, 1379 (1981); *J. Chem. Phys.* **74**, 1385 (1981)

[6] J E Howard, *Phys. Rev.* **A51**, 3934 (1995)
J Gao and J B Delos, *Phys. Rev.* **A49**, 869 (1994)

[7] M Courtney, N Spellmeyer, H Jiao and D Kleppner, *Phys. Rev.* **A51**, 3604 (1995)

[8] I Seipp and K T Taylor, *J. Phys.* **B27**, 2785 (1994)

[9] I Seipp, K T Taylor and W Schweizer, *J. Phys.* **B29**, 1 (1996)

[10] M Lakshmanan and K Ganesan, *Curr. Sci.* **68**, 38 (1995)
T S Monteiro, *Contemp. Phys.* **35**, 311 (1994)
J B Delos, S K Knudson and D W Noid, *Phys. Rev.* **A30**, 1208 (1984)

[11] C Cohen Tannoudji, B Diu and F Laloe, *Quantum Mechanics* (Wiley, New York, 1977)
L D Landau, E M Lifshitz, *Quantum mechanics—nonrelativistic theory* (Pergamon Press, Oxford, 1965)

[12] C W Clark and K T Taylor, *J. Phys.* **B13**, L737 (1980); *J. Phys.* **B15**, 1175 (1982); *Nature* **292**, 437 (1982)

[13] D Delande and A Buchleitner, *Adv. At. Mol. Opt. Phys.* **35**, 85 (1994)
J Wals, H H Fielding, J F Christian, L C Snoek, W J van der Zande and H B van Linden van den Heuvell, *Phys. Rev. Lett.* **72**, 3783 (1994)

[14] H Fredrich and D Wintgen, *Phys. Rep.* **183**, 37 (1989)
H Hasegawa, M Robnik and G Wunner, *Prog. Theor. Phys. Suppl.* **98**, 198 (1989)

[15] D R Herrick, *Phys. Rev.* **A26**, 323 (1982)
J B Delos, S K Knudson and D W Noid, *Phys. Rev.* **A28**, 7 (1983)

[16] H Fredrich and M C Chu, *Phys. Rev.* **A28**, 1423 (1983)

[17] W R S Garton and F S Tomkins, *Astrophys. J.* **158**, 839 (1969)

[18] A R P Rau, *J. Phys.* **B12**, L193 (1979)

[19] C Iu, G R Welch, M M Kash, D Kleppner, D Delande and J C Gay, *Phys. Rev. Lett.* **66**, 145 (1991)

[20] K T Taylor, M H Nayfeh and C W Clark (eds), *Physics of atoms and molecules: Atomic spectra and collisions in external fields* (Plenum Press, London, 1988)

[21] D Delande and J C Gay, *Phys. Rev. Lett.* **57**, 2006 (1986)
G Wunner, U Woelk, I Zech, G Zeller, T Ertl, F Geyer, W Schweitzer and H Ruder, *Phys. Rev. Lett.* **57**, 3261 (1986)

[22] D Delande, A Bommier and J C Gay, *Phys. Rev. Lett.* **66**, 141 (1991)
[23] J Zakrzewski, K Dupret and D Delande, *Phys. Rev. Lett.* **74**, 522 (1995)
K Dupret, J Zakrzewski and D Delande, *Europhys. Lett.* **31**, 251 (1995)
[24] P M Koch and K A H Van Leeuwen, *Phys. Rep.* **225**, 289 (1995)
[25] A D Peters, C Jaffe and J B Delos, *Phys. Rev. Lett.* **73**, 2825 (1994)
M A Iken, F Borondo, R M Benito and T Uzer, *Phys. Rev.* **A49**, 2734 (1994)
[26] P Cacciani, E Luc-koenig, J Pinard, C Thomas and S Liberman, *Phys. Rev. Lett.* **56**, 1467 (1986)
[27] H H Fielding, J Wals, W J van der Zande and H B van Linden van den Heuvell, *Phys. Rev.* **A51**, 611 (1995)
J M Amo, K A Rapelje, S J Blodgett-Ford, J B Delos, A Konig and H Rinneberg, *Phys. Rev.* **A48**, 2117 (1993)
[28] K Richter, D Wintgen and J S Briggs, *J. Phys.* **B20**, L627 (1987)
[29] K T Taylor, W S Schweizer, M A Al-Laithy, P F O'Mahony, S Bivone and C M Farmer, in: *Classical dynamics in atomic and molecular physics* (World Scientific, Singapore, 1989)
[30] M Courtney, *Phys. Rev.* **A51**, 4558 (1995)
[31] P Fassbinder and W Schweizer, *Phys. Rev.* **A53**, 2135 (1996)
Jan von Milczewski, G H F Diercksen and T Uzer, *Phys. Rev. Lett.* **76**, 2890 (1996)
[32] G Raithel, M Fauth and H Walther, *Phys. Rev.* **A44**, 1898 (1991)
[33] T Uzer and D Farrelly, *Phys. Rev.* **A52**, R2501 (1995)
[34] G Raithel and H Walther, *Phys. Rev.* **A49**, 1646 (1994)
G C Hegerfeldt and R Henneberg, *Phys. Rev.* **A49**, 3531 (1994)
Jan von Milczewski, G H F Diercksen and T Uzer, *Phys. Rev. Lett.* **73**, 2428 (1994)
[35] J Main and G Wunner, *J. Phys.* **B27**, 2835 (1994)
[36] K Ganesan and K T Taylor, *J. Phys.* **B29**, 1293 (1996)
[37] V S Melezhik, *Phys. Rev.* **A52**, R3393 (1995)
[38] K Ganesan and M Lakshmanan, *Phys. Rev. Lett.* **62**, 232 (1989); *Phys. Rev.* **A42**, 3940 (1990); *Phys. Rev.* **A45**, 1548 (1992); *Phys. Rev.* **A48**, 964 (1993); *J. Phys.* **B27**, 2809 (1994)
[39] J E Howard and D Farrelly, *Phys. Lett.* **A178**, 62 (1993)
D Farrelly and J E Howard, *Phys. Rev.* **A48**, 851 (1993)
G Baumann and T F Nonnenmacher, *Phys. Rev.* **A46**, 2682 (1992)
[40] M Lakshmanan and K Ganesan, *J. Phys. Soc. Jpn. Suppl.* **A63**, 47 (1994)
[41] R Blümel, C Kappler, W Quint and H Walther, *Phys. Rev.* **A40**, 808 (1989)
[42] R A Faulkner, *Phys. Rev.* **184**, 713 (1969)
[43] D Wintgen and H Marxer, *Phys. Rev. Lett.* **60**, 971 (1988)
[44] M C Gutzwiller, *Physica* **D5**, 183 (1992)
[45] D Wintgen, H Marxer and J S Briggs, *J. Phys.* **A20**, L965 (1987)
[46] André Bandrauk (ed.), *Atoms and molecular processes with short intense laser pulses*, NATO ASI Series B: Physics (Plenum Press, New York, 1988) vol. 171
[47] J E Bayfield and P M Koch, *Phys. Rev. Lett.* **33**, 258 (1974)
[48] T F Gallagher, *Mod. Phys. Lett.* **B5**, 259 (1991)
M Gatzke, B Broers, L D Noordam, R B Watkins and T Gallagher, *Phys. Rev.* **A50**, 2502 (1994)
R Blümel, A Buchleitner, R Graham, L Sirko, U Smilansky and H Walther, *Phys. Rev.* **A44**, 4521 (1991)
[49] M W Bellermann, P M Koch, D R Mariani and D Richards, *Phys. Rev. Lett.* **76**, 892 (1996)
[50] P Fu, T J Scholz, J M Hettema and T F Gallagher, *Phys. Rev. Lett.* **64**, 511 (1990)
[51] J G Leopold and I C Percival, *Phys. Rev. Lett.* **41**, 944 (1978)
D Banks and J G Leopold, *J. Phys.* **B11**, 37 (1978)
J G Leopold and I C Percival, *J. Phys.* **B12**, 709 (1979)
N B Delone, B A Zon and V P Krainov, *Zh. Eksp. Teor. Fiz.* **75**, 445 (1978); [*Sov. Phys. JETP* **48**, 223 (1978)]
B I Meerson, E A Oks and P V Sasorov, *Pisma Zh. Eksp. Teor. Fiz.* **29**, 79 (1979); [*Sov. Phys. JETP Lett.* **29**, 72 (1979)]
[52] K Burnett, V C Reed and P L Knight, *J. Phys.* **B26**, 561 (1993)

[53] A O Barut, C K E Schneider and R Wilson, *J. Math. Phys.* **20**, 2244 (1979)
A C Chen, *Phys. Rev.* **A22**, 333 (1980); **23**, 1655 (1981)
[54] R Gębarowski and J Zakrzewski, *Phys. Rev.* **A51**, 1508 (1995)
[55] M Gajda, J Grochmalicki, M Lewenstein and K Rzążewski, *Phys. Rev.* **A46**, 1638 (1992)
[56] R S MacKay, J D Meiss and I C Percival, *Phys. Rev. Lett.* **52**, 697 (1984)
[57] J D Meiss, *Phys. Rev. Lett.* **62**, 1576 (1989)
[58] C F Hillermeier, R Blümel and U Smilansky, *Phys. Rev.* **A45**, 3486 (1992)
[59] Y-C Lai, C Grebogi, R Blümel and M Ding, *Phys. Rev.* **A45**, 8284 (1992)
[60] J E Bayfield, S Y Luie, L C Perotti and M P Skrzypkowski, *Phys. Rev.* **A53**, R12 (1996)
[61] R Gębarowski and J Zakrzewski, *Phys. Rev.* **A50**, 4408 (1994)
[62] A Buchleitner, D Delande, J Zakrzewski, R N Mantegna, M Arndt and H Walther, *Phys. Rev. Lett.* **75**, 3818 (1995)
[63] K Kulander, K J Schafer and J L Krause, *Phys. Rev. Lett.* **66**, 2601 (1991)
J L Krause, K J Schafer and K Kulander, *Phys. Rev.* **A45**, 4998 (1992)
[64] J Parker, K T Taylor, C W Clark and S Blodgett-Ford, *J. Phys.* **B29**, L33–42 (1996)
[65] J H Shirley, *Phys. Rev.* **B138**, 979 (1965)
S-I Chu and W P Reinhardt, *Phys. Rev. Lett.* **39**, 1195 (1977)
[66] A Buchleitner, D Delande and J-C Gay, *J. Opt. Soc. Am.* **B12**, 505 (1995)
[67] J Zakrzewski, D Delande, J C Gay and K Rzążewski, Phys. Rev. **A47**, R2468 (1993)
[68] B V Chirikov, *Phys. Rep.* **52**, 263 (1979)
[69] M V Berry and K E Mount, *Rep. Prog. Phys.* **35**, 315 (1972)
C D Schwieters and J B Delos, *Phys. Rev.* **A51**, 1023 (1995)
[70] F Benvenuto, G Casati and D L Shepelyansky, *Phys. Rev.* **A45**, R7670 (1992); *Phys. Rev.* **A47**, R786 (1993)
[71] G Casati, I Guarneri and D L Shepelyansky, *IEEE J. Quant. Electron.* **24**, 1240 (1988)
[72] M Wójcik, J Zakrzewski and K Rzążewski, *Phys. Rev.* **A52**, R2523 (1995)
[73] J Zakrzewski and K Życzkowski, *Phys. Rev.* **A36**, 4311 (1987)
[74] D Wintgen, *Z. Phys.* **D18**, 125 (1991)
[75] M J Raković and S-I Chu, *Phys. Rev.* **A50**, 5077 (1994)
[76] K Rzążewski and B Piraux, *Phys. Rev.* **A47**, R1612 (1993)
[77] J Mostowski and J J Sanchez-Mondragon, *Opt. Commun.* **29**, 293 (1979)
B I Meerson, E A Oks and P V Sasorov, *J. Phys.* **B15**, 3599 (1982)
[78] J Zakrzewski, R Gębarowski and D Delande, *Phys. Rev.* **A54**, 691 (1996)
[79] M Nauenberg, *Europhys. Lett.* **13**, 611 (1990)
[80] J E Howard, *Phys. Rev.* **A46**, 364 (1992)
[81] P Kappertz and M Nauenberg, *Phys. Rev.* **A47**, 4749 (1993)
[82] D Farrelly and T Uzer, *Phys. Rev. Lett.* **74**, 1720 (1995)
[83] I Białynicki-Birula, M Kaliński and J H Eberly, *Phys. Rev. Lett.* **73**, 1777 (1994)
M Kaliński, J H Eberly and I Białynicki-Birula, *Phys. Rev.* **A52**, 2460 (1995)
[84] D Delande, J Zakrzewski and A Buchleitner, *Europhys. Lett.* **32**, 107 (1995)
[85] J Zakrzewski, D Delande and A Buchleitner, *Phys. Rev. Lett.* **75**, 4015 (1995)
[86] G Casati, B V Chirikov, D L Shepelyansky and I Guarneri, *Phys. Rep.* **154**, 77 (1987)
[87] S Benkadda, Y Elskens, B Ragot and J T Mendonça, *Phys. Rev. Lett.* **72**, 2859 (1994)
[88] R Blümel and U Smilansky, *Z. Phys.* **D6**, 83 (1987)
[89] G Hose and H S Taylor, *Phys. Rev. Lett.* **51**, 947 (1983)
[90] J Zakrzewski, D Delande and M Kuś, *Phys. Rev.* **E47**, 1665 (1993)

—journal of
physics

February 1997
pp. 411–424

Resonances and chaos in the collinear collision system $(\mathrm{He}, \mathrm{H}_2^+)$ and its isotopic variants

SUSANTA MAHAPATRA[a], N SATHYAMURTHY[a] and
RAMAKRISHNA RAMASWAMY[b]
[a] Department of Chemistry, Indian Institute of Technology, Kanpur 208 016, India
[b] School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Abstract. The collinear atom-diatom collision system provides one of the simplest instances of chaotic or irregular scattering. Classically, irregular scattering is manifest in the sensitive dependence of post-collision variables on initial conditions, and quantally, in the appearance of a dense spectrum of dynamical resonances. We examine the influence of kinematic factors on such dynamical resonances in collinear $(\mathrm{He}, \mathrm{H}_2^+)$ collisions by computing the transition state spectra for collinear $(\mathrm{He}, \mathrm{HD}^+)$ and $(\mathrm{He}, \mathrm{DH}^+)$ collisions using the time-dependent quantum mechanical approach. The nearest neighbor spacing distribution $P(s)$ and the spectral rigidity $\Delta_3(L)$ for these resonances suggest that the dynamics is predominantly *irregular* for collinear $(\mathrm{He}, \mathrm{HD}^+)$ and predominantly *regular* for collinear $(\mathrm{He}, \mathrm{DH}^+)$. These findings are reinforced by a significantly larger "correlation hole" in ensemble averaged survival probability $\langle\langle P(t)\rangle\rangle$ values for collinear $(\mathrm{He}, \mathrm{HD}^+)$ than for collinear $(\mathrm{He}, \mathrm{DH}^+)$. In addition we have also examined measures of classical chaos through the dependence of the final vibrational action, n_f, on the initial vibrational phase ϕ_i of the diatom, and Poincaré surfaces–of–section. They show that $(\mathrm{He}, \mathrm{HD}^+)$ collisions are partly chaotic over the entire energy range (0–2.78 eV) while $(\mathrm{He}, \mathrm{DH}^+)$ collisions, in contrast, are highly regular at collision energies below the classical threshold for reaction. Above the threshold, the scattering remains regular for initial vibrational states $v = 0$ and 1 of DH^+.

Keywords. Dynamical resonances; irregular scattering; collinear collisions.

PACS Nos 05.45; 34.10; 34.50

1. Introduction

Dynamical resonances in atom-molecule collisions [1] have been of great interest since they give rise to unusual variations in the scattering amplitude and related quantities. Reactive scattering resonances, which have been identified in a number of systems, one of which is $(\mathrm{He}, \mathrm{H}_2^+)$ in collinear as well as noncollinear configurations [2], are characterized by oscillations in the reaction probability P^R as a function of the collision energy E. Many of them are of the Feshbach type and can be interpreted in terms of bound states supported by vibrational adiabatic potentials in hyperspherical coordinates [2(d),3]. A closer examination of $P^R(E)$ curves for collinear $(\mathrm{He}, \mathrm{H}_2^+)$, for example, reveals that there are additional oscillations [2(h)] that are irregularly spaced as a function of the energy and which can not be assigned easily. The connection between such (quantal) resonances and the features of the classical dynamics is the subject of this article.

$$\mathrm{He} + \mathrm{H}_2^+ \rightarrow \mathrm{HeH}^+ + \mathrm{H} \qquad (\mathrm{R1})$$

reaction on the McLaughlin-Thompson-Joseph-Sathyamurthy (MTJS) potential energy surface (PES) [5] revealed that plots of final vibrational action (n_f) of the reactant/product molecule as a function of the initial vibrational phase (ϕ_i) of the reactant molecule consists of reactive (R) and nonreactive (NR) bands and a number of "chattering" or irregular trajectories in between. Upon magnification, it was found that the primary chattering region actually consists of a number of alternating R and NR bands with secondary chattering regions separating them. On further magnification, it became evident that there is a self-similar pattern repeating itself endlessly on increased resolution: $n_f(\phi_i)$ is a fractal [6] curve. Similar behavior has been found for several other systems like collinear $(\mathrm{H}, \mathrm{H}_2)$ [7] and (Li, FH) [8]. Trajectories in the chattering regions tend to be chaotic, and the origin of classical chaos in collinear $(\mathrm{He}, \mathrm{H}_2^+)$ system has been examined [9,10] by plotting the Poincaré surface of section (SOS) for a number of trajectories at energies below the reaction threshold.

It has been shown from time-dependent quantal wave packet (WP) calculations [2(h)] that the dynamics in collinear $(\mathrm{He}, \mathrm{H}_2^+)$ collisions is "turbulent" when viewed in terms of probability density plots in configuration space. The transition state spectrum [11] is obtained by time evolving a wave packet centered in the interaction region, followed by computation of the Fourier transform of the autocorrelation function. Peaks in this spectrum correspond to dynamical resonances, and statistical analysis of the resonance energies, through the nearest neighbor spacing distribution (NNSD) or the spectral rigidity revealed an intermediate (partly regular and partly irregular) behavior for the system [12]. An ensemble averaged survival probability $\langle\langle P(t)\rangle\rangle$, computed [12] by Fourier transforming the spectral autocorrelation function to time domain followed by averaging over the initial states and Hamiltonian revealed the existence of a "correlation hole" that is graphic evidence for "irregularity" or quantum chaos [13–16]. A one-to-one correspondence between resonant periodic orbits (RPOs) and many transition state resonance eigenfunctions for the system also has been established [11, 17]. Therefore, on the whole, it could be concluded that there is a strong correlation between classical chaos, quantum chaos and dynamical resonances for systems like collinear $(\mathrm{He}, \mathrm{H}_2^+)$.

Since dynamical resonances are intimately connected with the nature of the motion, it may be anticipated that kinematic factors play a crucial role in such systems. In the present study we report on the spectral characteristics and the behavior of quasiclassical trajectories for the two widely differing dynamical systems which arise from isotopic substitution of HeH_2^+, in order to investigate the role of kinematic factors in dynamical resonances. Within the Born-Oppenheimer approximation, isotopic substitution has no influence on the PES. Therefore, any difference in dynamical results must arise from the difference in masses of the atoms concerned.

We have computed the $P_v^R(E)$ curves for collinear $\mathrm{He} + \mathrm{HD}^+$ and $\mathrm{He} + \mathrm{DH}^+$ collisions, with HD^+ in different initial vibrational states, v, and find that $P_v^R(E)$ in either case behaves very differently [18] (from each other as well as from the case of HeH_2^+). Our results suggest that the wild oscillations in $P_v^R(E)$ may not necessarily mean that the

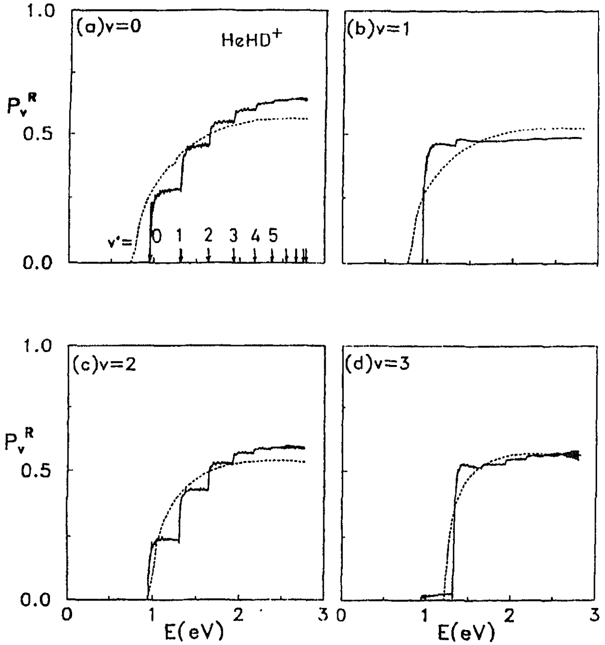


Figure 1. State-selected reaction probabilities (solid curve) computed from time-dependent quantal wave packet calculations and reproduced from Ref. [18] for different initial vibrational (v) states of HD^+ for the reaction $He + HD^+(v) \rightarrow HeH^+ +$ D. The corresponding quasiclassical trajectory results are included as dashed curves. The threshold energies of various vibrational channels of the product HeH^+ are marked along the abscissa of panel (**a**) to illustrate the occurrence of sharp variations in P_v^R near each channel threshold.

dynamics is chaotic from a quantal point of view. Collinear (He, HD^+) would correspond to a (**H, LH**) mass combination, while (He, DH^+) would, to a (**H, HL**) mass combination. Here **H** and **L** refer to heavy and light atoms, respectively. Based on the known $P_v^R(E)$ results for other systems of (**H, LH**) type, it is expected that (He, HD^+) would have an oscillatory $P_v^R(E)$ as a light atom is exchanged between two heavy atoms, and that (He, DH^+) would have a relatively structureless $P_v^R(E)$ since the leaving atom is light. However, we find that $P_v^R(E)$ for collinear (He, HD^+) has a staircase-like structure, with minor oscillations about each "step", as illustrated in figure 1 for different initial v states of HD^+. The sharp variations in $P_v^R(E)$ (from one step to another) can be readily identified [18] as threshold resonances [19] corresponding to the opening up of either reactant or product vibrational channels. The undulations of smaller magnitude about the steps are due to the weak dynamical resonances found in the transition state spectrum. The $P_v^R(E)$ curve for collinear (He, DH^+) collisions, on the other hand, is highly oscillatory as shown in figure 2 and such an interpretation would not suffice to explain the structure. We therefore examine the transition state spectra of both systems for signatures of quantum chaos, and find that collinear (He, HD^+) collisions are predominantly chaotic, whereas (He, DH^+) collisions

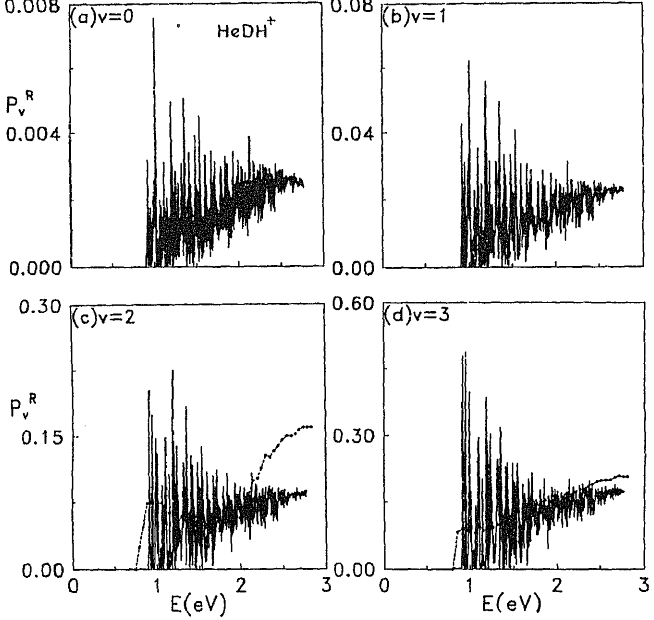


Figure 2. State-selected quantal reaction probabilities (solid curve) reproduced from Ref. [18] for the reaction $\mathrm{He} + \mathrm{DH}^+(v) \rightarrow \mathrm{HeD}^+ + \mathrm{H}$. The corresponding quasiclassical trajectory results are included as dashed curves. For $v = 0$ and 1, the quasiclassical trajectory calculations yield $P_v^R = 0$ (the statistical uncertainty level is 0.001).

are not. From a classical point of view, however, both systems show characteristics of chaos, albeit to different degrees.

2. Quantum chaos

We obtain the transition state spectra [18] for collinear $(\mathrm{He}, \mathrm{HD}^+)$ and $(\mathrm{He}, \mathrm{DH}^+)$ collisions by evolving a wave packet $\psi(t = 0)$ initially centered in the transition state region and computing the autocorrelation function $C(t) = \langle\psi(0)|\psi(t)\rangle$, and its Fourier transform. There are a large number of quasibound states, i.e. transition state resonances for both systems. These resonance energies $\{E_i, i = 1, 2 \ldots N\}$ provide the data for our subsequent analysis. Over the past couple of decades, considerable evidence has accumulated to support the hypothesis that the statistical properties of the eigenvalues of systems that have chaotic classical limits are identical to those of ensembles of random matrices [20] of the appropriate symmetry or universality class. Here we apply such analysis to the systems at hand.

The cumulative level density $N(E)$, which is a staircase function, is first unfolded in order to make the average density of energy levels uniform over the entire energy range. There are no clear-cut methods to unfold the spectrum of resonance states. Therefore we have adopted the procedure that is followed for nuclear resonances and bound states, namely that the unfolded energy levels $\bar{E}_i$ are calculated from the energy levels in the raw spectrum (E_i) by dividing all the level spacings by the local average spacing [21,22]

$$\begin{aligned} \bar{E}_{i+1} = & \bar{E}_i + (2k+1)\frac{E_{i+1}-E_i}{E_{j_2+1}-E_{j_1}}; \\ & j_1 = i-k, (\overline{i-k} > 0); \quad j_2 = i+k, (\overline{i+k} < n), \end{aligned} \tag{1}$$

where k is the number of consecutive spacings used to calculate the local average spacing and n is the total number of energy levels.

From the unfolded spectrum, the nearest neighbor spacing distribution, $P(s)$ can be obtained; the spacings are $\{s_i = \bar{E}_{i+1} - \bar{E}_i, i = 1, 2 \ldots N-1\}$. In the limiting cases of purely chaotic or purely regular systems, the distributions that obtain are well known: the former is (to a high degree of accuracy) given by the Wigner surmise, $P(s) = (\pi s/2)\exp(-(\pi s^2/4))$ [23], while the latter is the Poisson distribution $P(s) = \exp(-s)$ [24].

The spectral rigidity $\Delta_3(L)$, computed from the unfolded spectrum measures the least squares deviation of the staircase function ($N(E)$) from the best straight line in the energy interval of length L [25] and it gives information on the long-range correlations—how one state influences another that is well separated from it. $\Delta_3(L)$ is computed from the unfolded energy values lying in the interval $[0, \ldots, L]$ as [26]

$$\begin{aligned} \Delta_3(L) = \Bigg\langle \left(\frac{n^2}{16}\right) - \frac{1}{4L}\left(\sum_{i=1}^{n} \bar{E}_i\right)^2 + \frac{3}{8L^2}\left(\sum_{i=1}^{n} \bar{E}_i\right)^2 - \frac{3}{16L^4}\left(\sum_{i=1}^{n} \bar{E}_i{}^2\right)^2 \\ + \frac{1}{2L}\left(\sum_{i=1}^{n}(n-2i+1)\bar{E}_i\right)\Bigg\rangle, \end{aligned} \tag{2}$$

where n is the number of levels within the energy interval L and the ensemble average $\langle\rangle$ is carried out by shifting the box of length L over the entire spectrum. The limiting behaviours are again well-known, and are

$$\Delta_3^{\text{Poisson}}(L) = L/15 \tag{3}$$

and

$$\Delta_3^{\text{GOE}}(L) \approx \frac{1}{\pi^2}\log L - 0.00695 \tag{4}$$

respectively for the regular and chaotic cases [20,23,27–29], the superscript GOE referring to the Gaussian Orthogonal Ensemble of random matrices [20,30]. Molecular systems are often neither purely regular nor purely chaotic, in which case the mixed distribution [31,32] applies,

$$\Delta_3(L; q) = \Delta_3^{\text{Poisson}}((1-q)L) + \Delta_3^{\text{GOE}}(qL), \tag{5}$$

where q represents the fraction of phase space that is chaotic.

For HeHD$^+$ the NNSD distribution, shown in figure 3(a), is more similar to the chaotic case than the regular, indicative of level repulsion. This is confirmed by the behaviour of the $\Delta_3(L)$ statistic, shown in figure 3(b): shown along with the data (triangles) for HeHD$^+$ are the GOE and Poisson values. Except at small L, the computed points for

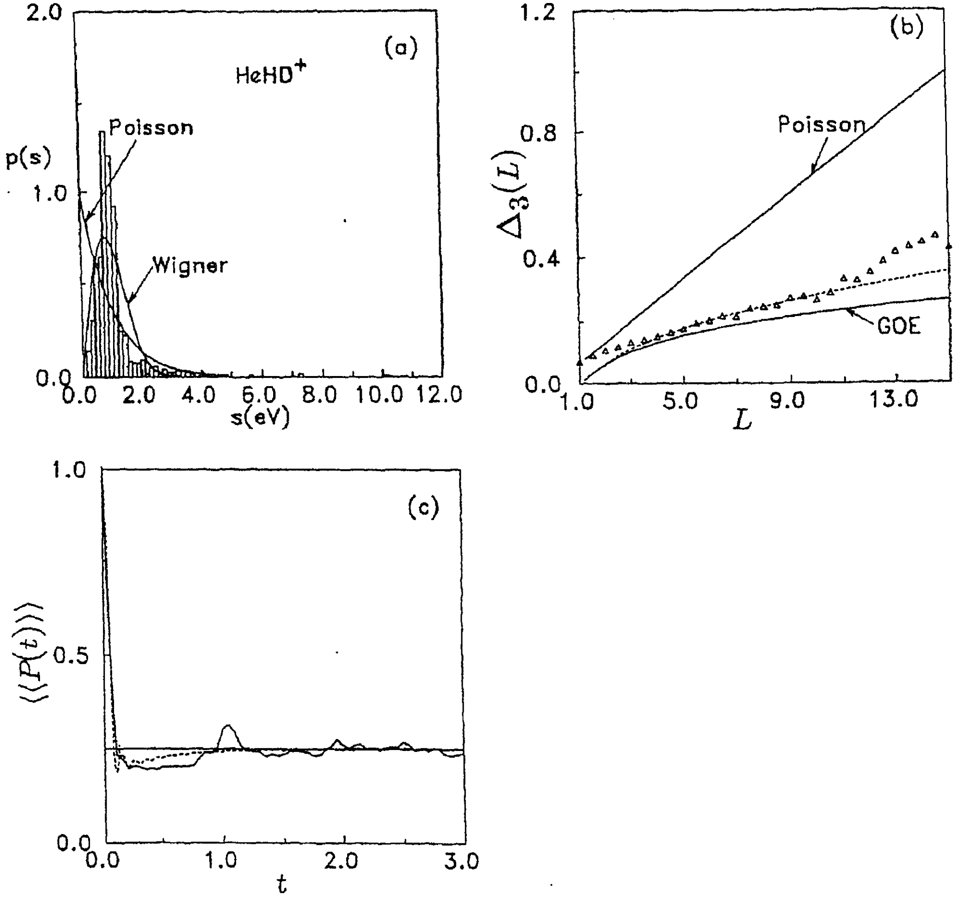


Figure 3. **(a)** The nearest neighbor spacing distribution $P(s)$ derived from the unfolded eigenvalue spectrum for collinear HeHD$^+$, showing short-range correlations. Poisson and Wigner indicate the distributions in the regular and the irregular limits, respectively. **(b)** The spectral rigidity $\Delta_3(L)$ indicating the long-range correlations as a function of the segment length, L, in the above spectrum (triangular points). $\Delta_3(L)$ computed from Eq. 5 for mixed behavior, using $q = 0.9$ is shown by the dashed curve. Poisson and GOE indicate the regular and the irregular limits respectively. **(c)** Variation of ensemble averaged survival probability $\langle\langle P(t)\rangle\rangle$ with time t (in arbitrary units) obtained from the above spectrum (solid curve). The RMT $\langle\langle P(t)\rangle\rangle$ values for $\beta = 0.9$ are included as the dashed curve, for comparison.

HeHD$^+$ are closer to the GOE limit than to the Poisson, thus revealing a high degree of rigidity in the spectrum, more so than the case of collinear (He,H$_2^+$). Since the computed $\Delta_3(L)$ distribution does not fully correspond to either the Poisson or the GOE value, we have fitted it to the intermediate case (Eq. 5), and obtain the dashed line in figure 3(b), using $q = 0.90$.

For collinear HeDH$^+$ on the other hand, the NNSD in figure 4(a) is more ambiguous: it is difficult to decipher whether it is Poisson-like or Wigner-like. The $\Delta_3(L)$ values

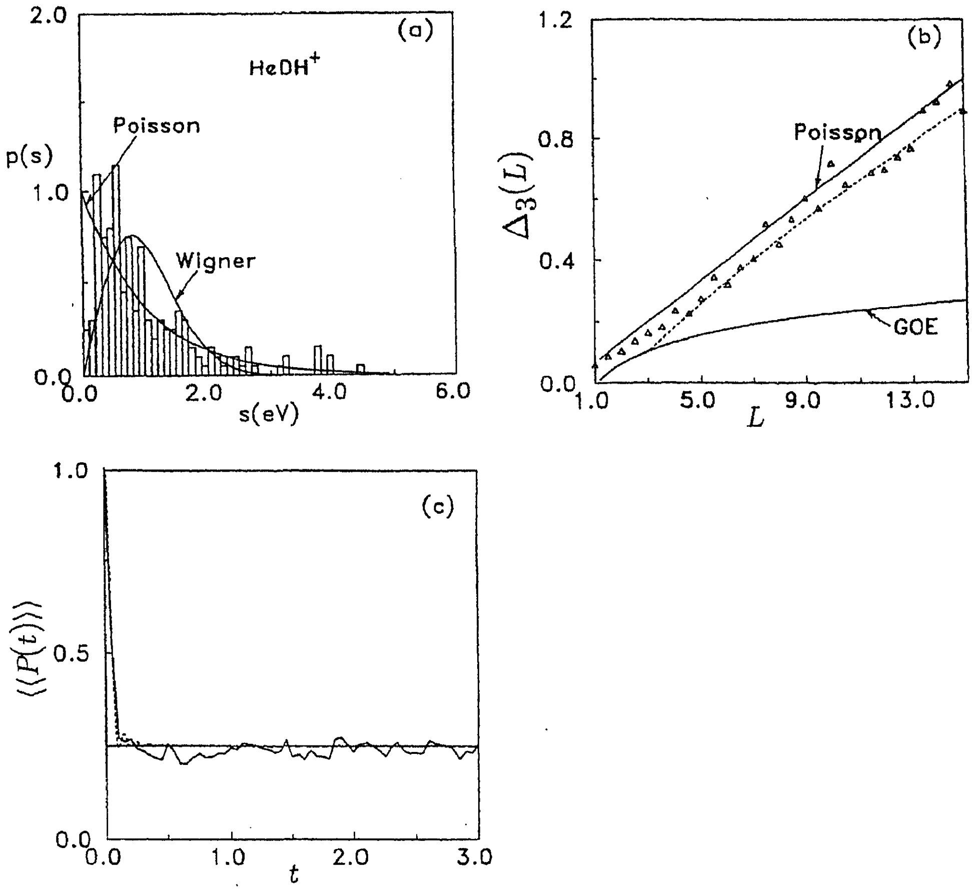


Figure 4. (**a**) The nearest neighbor spacing distribution derived from the unfolded eigenvalue spectrum for collinear HeDH$^+$. (**b**) The spectral rigidity for the same spectrum (triangular points). The dashed curve corresponds to $\Delta_3(L)$ values obtained from Eq. 5 for $q = 0.2$. (**c**) The ensemble averaged survival probability with time t (in arbitrary units), obtained from the above spectrum (solid curve). The RMT$\langle\langle P(t)\rangle\rangle$ values for $\beta = 0.2$ are included for comparison (the dashed curve).

obtained from the unfolded spectrum, shown in figure 4(b) as triangular points, are close to the Poisson distribution. These values are best fitted to Eq. 5 (shown as the dashed curve) using $q = 0.2$, which indicates, in contrast to HeHD$^+$, a high degree of regular behavior.

Further analysis of the dynamical resonance spectra is possible. In the time-domain, analysis of the transition state spectrum proceeds through the computation of the ensemble averaged survival probability $\langle\langle P(t)\rangle\rangle$ [13–16]. The survival probability, averaged over the initial states $\langle P(t)\rangle$, is computed [16] through

$$\langle P(t)\rangle = \frac{3}{N+2}\left\{1 + \frac{2}{3N}\sum_{n,m;n>m}\cos\left[2\pi(\bar{E}_n - \bar{E}_m)\frac{t}{\langle\Delta\bar{E}\rangle}\right]\right\} \tag{6}$$

and further averaging over the Hamiltonian is done by averaging over n levels which are divided into several segments, each containing N levels, to obtain $\langle\langle P(t)\rangle\rangle$. $\langle\Delta\bar{E}\rangle$ is the average level spacing. The $\langle\langle P(t)\rangle\rangle$ values thus obtained from the unfolded spectrum of collinear HeHD^+ are shown by the solid curve in figure 3(c). The dip below the asymptotic value at short times ($t \approx 2\pi\hbar/\langle\Delta\bar{E}\rangle$), termed a "correlation hole" is a typical signature of irregularity in the spectrum. The extent of irregularity is estimated by comparing the $\langle\langle P(t)\rangle\rangle$ values with those derived from random matrix theory (RMT) [20]. The latter are obtained from [33]

$$\langle\langle P(t)\rangle\rangle = \frac{3}{N+2}\{1 + \tfrac{1}{3}\Delta_N * [\delta(\tau) - b_{2\beta}(\tau)]\}, \tag{7}$$

where $\tau = t/(2\pi\langle\rho\rangle)\langle\rho\rangle$ is the average level density) and $\beta = 1$, 2 and 4 for Gaussian orthogonal, unitary and symplectic ensembles, respectively. $b_{2\beta}(t)$ is the two-level form factor obtained by Fourier transformation of the two-point cluster function, $Y_{2\beta}(\omega)$ [20]. The asterisk implies convolution and $\Delta_N(t) = N^{-1}[\sin(\pi N t)/\pi t]^2$. If fraction β of the levels obey GOE statistics and (1-β) obey Poisson statistics, $b_{2\beta}(\mathrm{t})$ ($0 \leq \beta \leq 1$) is given by [33]

$$b_{2\beta}(t) = \beta b_2(t/\beta) + [(1-\beta)/N]\delta(t) \tag{8}$$

where b_2 indicates the GOE value that is obtained from the two-level cluster function as [20]

$$\begin{aligned} b_2(t) &= 1 - 2|t| + |t|\log(1 + 2|t|), && |t| \leq 1, \\ &= -1 + |t|\log((2|t| + 1)/(2|t| - 1)), && |t| \geq 1. \end{aligned} \tag{9}$$

For HeHD^+, the $\langle\langle P(t)\rangle\rangle$ values obtained from Eq. 7 for $\beta = 0.9$ (dashed line) are found to be in reasonable agreement with those derived from the spectrum through Eq. 6 (the solid line) as illustrated in figure 3(c). We conclude that approximately 90% of the resonance spectrum for HeHD^+ is irregular in nature. This estimate for the fractional irregularity of the eigenvalue spectrum is also in agreement with the estimate obtained from $\Delta_3(L)$ (see above).

For HeDH^+ on the other hand, the $\langle\langle P(t)\rangle\rangle$ curve, shown in figure 4(c) reveals only a marginal correlation hole (if any) implying that collinear ($\mathrm{He}, \mathrm{DH}^+$) dynamics is highly regular. In addition, the above $\langle\langle P(t)\rangle\rangle$ values, when compared to those obtained from Eq. 7, (the dashed line) yield a value of $\beta = 0.2$, which is consistent with the fractional irregularity of the spectrum obtained from $\Delta_3(L)$.

3. Classical chaos

In order to see how the classical dynamics of the two systems differ we have computed quasiclassical trajectories for both over a wide range of energies. The resulting $P_v^R(E)$ values are included in figures 1 and 2 in the form of dashed lines. As is often the case, the classical threshold (E_{th}^{cl}) values for the exchange reaction in the case of $\mathrm{He} + \mathrm{HD}^+ (v = 0, 1)$ are lower than the quantal (E_{th}^{qm}) values because products with $n_f < 0$ are possible in classical mechanics [34] and are indeed found to be formed. For $v = 2$, $E_{th}^{cl} = E_{th}^{qm}$. For $v = 3$, $E_{th}^{cl} > E_{th}^{qm}$ as the exchange reaction is dynamically

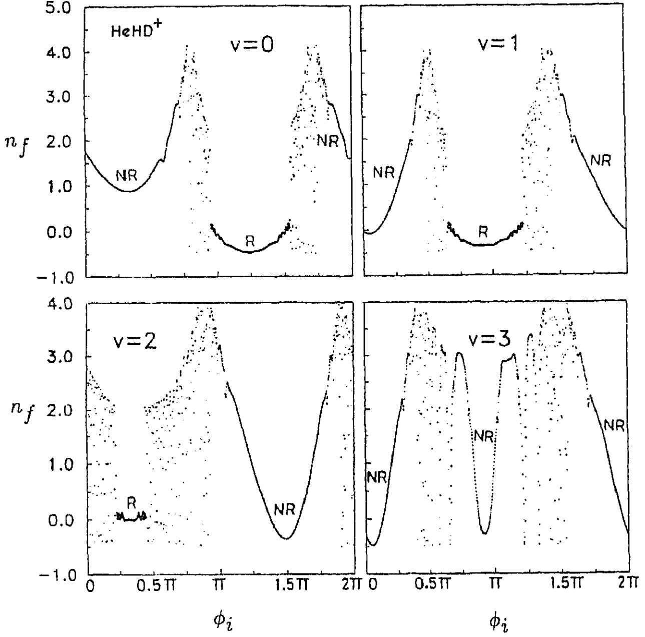


Figure 5. Action-angle plots for a family of 1000 trajectories for various initial vibrational (v) states (indicated in each panel) of HD^+ in collinear He, $HD^+(v)$ collisions at $E_{tot} = 1.0$ eV. R and NR stand for the reactive and the nonreactive band of trajectories, respectively.

forbidden in the vicinity of E_{th}^{qm}. For He + $DH^+(v = 0, 1)$, classical mechanics predicts $P_v^R = 0$ while quantum mechanics yields extremely low (but nonzero) reaction probabilities, P_v^R. For $v = 2$ and 3, $E_{th}^{cl} < E_{th}^{qm}$ as was discussed above. We note, parenthetically, that it is clear that on average, classical mechanics gives a good description of collinear He, HD^+ and He, DH^+ dynamics. The only major discrepancy (over and above the threshold effect) between the classical and quantal results is seen to occur for He, $DH^+(v = 2)$ at high energies and its origin is not clear.

Scattering chaos is most evident [35] when the final quantum number (related to the final classical action) n_f is plotted versus the initial phase, ϕ_i, as shown in figures 5 and 6 for collinear (He, HD^+) and (He, DH^+) collisions for different v states of HD^+ at a total energy (E_{tot}) of 1.0 eV. It is clear from figures 5(a) and 5(b) that there are R and NR bands and chattering regions in between, for (He, $HD^+(v = 0, 1)$). For $v = 2$, the R band has shrunk in size and the chattering region has become wider, when compared to that for $v = 0$ and 1. For $v = 3$, the R band has disappeared completely. There are only NR bands and chattering regions in between. It can be seen that there is a prominent parabola or *icicle* at the centre and there are narrower icicles on either side [36]. A larger number of trajectories computed in the chattering region would reveal fractal characteristics. Since (He, DH^+) collisions are nonreactive for $DH^+(v = 0, 1)$ we obtain only an NR band with slight distortions, as can be seen in figures 6(a) and 6(b). For $v = 2$ and 3 we obtain an NR band and a set of reactive trajectories whose n_f values vary rapidly with ϕ_i. A

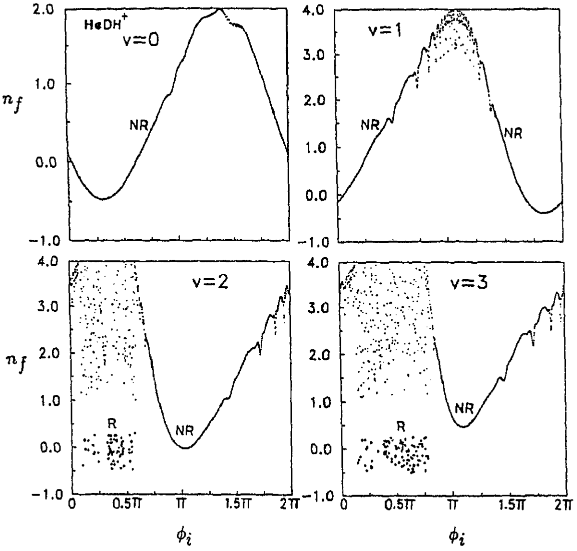


Figure 6. Same as figure 5 for He, DH$^+(v)$ collisions.

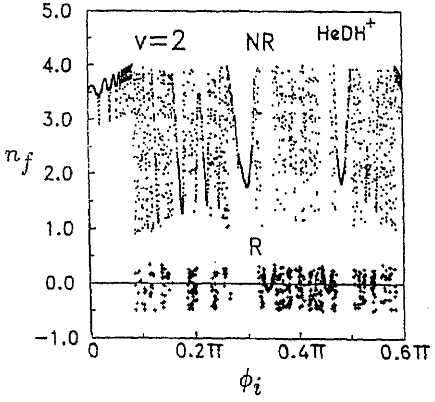


Figure 7. Action-angle plot at increased resolution for He, DH$^+(v = 2)$ collisions. R and NR represent the reactive and the nonreactive trajectories, respectively. The root trajectories that determine the scattering amplitude in the classical S-matrix theory correspond to the intersection of the fractal curve (R) with the line $n_f = 0$; it is evident that there are a very large number (essentially infinite) of these.

blow-up of the chattering region containing reactive trajectories is shown in figure 7. The reactive and the nonreactive trajectories are interspersed along the ϕ_i axis. But because of the differing range of n_f values for the two types of collisions, they are well separated along the vertical axis, but both reactive as well as nonreactive trajectories have characteristic icicle structure. From the point of view of semiclassical S-matrix theory [37], the horizontal line at $n_f = 0$ implies that there is a large number of "root" trajectories and that the expression for the reaction probability would consist of a large

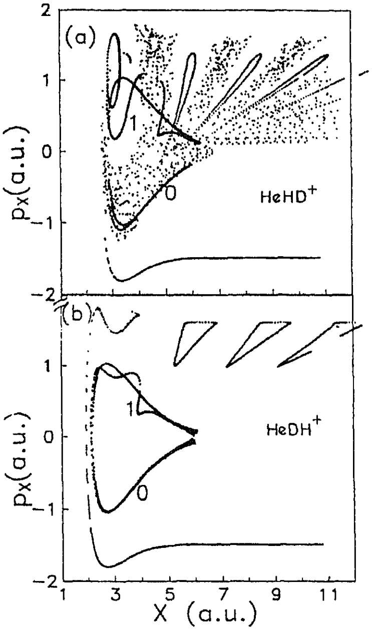


Figure 8. Poincaré surface of section obtained from 400 trajectories superimposed on the zero (0) and the first (1) order separatrices at $E_{tot} = 0.6434$ eV for collinear (**a**) He, HD$^+$($v = 0$) and (**b**) He, DH$^+$($v = 0$) collisions.

number of interference terms, thus explaining the highly oscillatory nature of $P_v^R(E)$ curve for the reaction.

In order to investigate the classical chaos in collinear (He, HD$^+$) and (He, DH$^+$) collisions further, we show in figure 8(a) plots of Poincaré SOS for 400 trajectories with HD$^+$($v = 0$) at E_{tot}=0.6434 eV, along with the zero and the first order separatrices. The zero order separatrix is obtained by solving the implicit equation

$$\frac{p_x^2}{2M} + V(x, y = y_e) = 0, \tag{10}$$

where M is the atom-diatom reduced mass and $V(x, y = y_e)$ is the potential energy of the system, with the diatom in its equilibrium configuration. The first order separatrix is obtained by propagating several sets of (x, p_x) values on the zero order separatrix forward in time by setting $y = y_e$ and $p_y = \sqrt{2mE_v}$, where m is the diatomic reduced mass and E_v is the vibrational energy of the diatom, equal to E_{tot}, until it crosses the SOS again. The first order separatrix, when superimposed on the zero order separatrix, reveals a turnstile or broken separatrix [38]. The latter consists of flux-in and flux-out regions. Direct (regular) trajectories never enter the intramolecular bottleneck (identified with the zero order separatrix) whereas indirect (irregular) trajectories do, invariably entering through the flux-in region and exiting through the flux-out region.

The results in figure 8(a) are very similar to those in figure 3 of Ref. [10] for collinear (He, H$_2^+$) . The regular trajectories never enter the intramolecular bottleneck and they distribution in the final relative translational energy. On the other hand the irregular trajectories enter largely through the broken separatrix and they undergo multiple collisions inside the strong intramolecular bottleneck region and form the complementary dusty pattern between the baseball bats.

The zero and the first order separatrices, along with the Poincaré SOS for a family of regularly sampled 400 trajectories for collinear (He, $DH^+(v = 0)$) at E_{tot} =0.6434 eV are shown in figure 8(b). Again, in contrast to figure 8(a), it is clear that all the trajectories are directly scattered. They never enter the intramolecular bottleneck, emphasizing the fact that collinear (He, DH^+) trajectories are predominantly regular at energies below E^{cl}_{th}. Even for energies above E^{cl}_{th}, the Poincaré SOS plots confirm that there are regular and irregular trajectories for collinear $\mathrm{He}, \mathrm{HD}^+(v = 0-3)$. For $\mathrm{He}, \mathrm{DH}^+(v = 0, 1)$ the trajectories are nonreactive but entirely regular. For $v = 2$ and 3 however, both regular and irregular trajectories occur.

4. Summary and conclusions

Chaotic scattering is sensitive to kinematic effects, and with this in mind, in the present article we have examined the transition state spectra of collinear (He,HD^+) and (He, DH^+) systems obtained from time-dependent quantal wave packet calculations. Signatures of quantum chaos become evident in statistical analyses of spacing distributions, the spectral rigidity and ensemble averaged survival probability. In this regard, the two isotopic variants we have looked appear to be dynamically very different, although the interaction potential is identical in both the cases.

The 3-body system (He, H_2^+) has the mass combination Heavy-Light-Light. Changing the terminal atom from H to D, namely by making this a Heavy-Light-Heavy system keeps the dynamics irregular, both classically as well as quantum mechanically. Changing the central atom to D gives the mass combination Heavy-Heavy-Light, and this dramatically reduces the chaotic behaviour: the $HeDH^+$ spectrum is regular despite the wild oscillatory $P^R_v(E)$ curve. Action-angle plots and Poincaré SOS plots indicate that (He, HD^+) dynamics is mixed i.e. partly regular and partly irregular, over the entire energy range of investigation. Results for (He, DH^+), in contrast, suggest that the dynamics is nearly completely regular for $v = 0$ and 1, but for $v = 2$ and 3, it assumes a mixed character.

As was mentioned in the Introduction, classical chaos and quantum mechanical resonances have been investigated in collinear (H, H_2)[7], (Li, FH) [8] and other systems. Quantal resonances have been reported for three dimensional (He, H_2^+) [2e–2g] and (Li, FH) [39] collisions as well. The large number of spectral lines observed in the infrared predissociation spectrum of H_3^+ [40] have been traced [41] to bound states embedded in the continuum (read transition state resonances) of the H_2–H^+channel and also related to horse-shoe periodic orbits. Indications of the role of kinematic factors in the system came from the D_2H^+ spectra that depended on the formation of H^+/D^+.

Attempts have been made [42] in the past to verify experimentally the existence of resonances in (He, H_2^+) collisions. Following our suggestion [2d] that observables like state-to-state differential cross sections for backward scattering, which are decided largely by collinear collisions at zero impact parameter, might be amenable to experimental verification, Pollard *et al* [43] measured the same. Unfortunately, their energy resolution (0.15 eV) was not good enough to detect the narrow resonances in the system. Since $P_0^R(E)$ behaves differently for HeH^+ and HeD^+ formation in collinear collisions, it is quite likely that this effect would persist in three dimensions. In that case, measurement of differential cross sections for the two different channels in (He, HD^+) collisions might reveal some of the interesting features reported in this paper.

Acknowledgement

This study was supported in part by a grant from the Commission of European Communities.

References

[1] (a) D G Truhlar (Ed.), *Resonances in electron-molecular scattering, van der Waals complexes and reactive chemical dynamics*, ACS symposium series No. 263, American chemical society, Washington, D. C., 1984
  (b) G C Schatz, *Ann. Rev. Phys. Chem.* **39**, 317 (1988)
  (c) D E Manolopoulos and D C Clary, *Ann. Rep. Chem. Soc.* **C95**, (1989)
  (d) W H Miller, *Ann. Rev. Phys. Chem.* **41**, 245 (1990)

[2] (a) D J Kouri and M Baer, *Chem. Phys. Lett.* **24**, 37 (1974)
  (b) J T Adams, *Chem. Phys. Lett.* **33**, 275 (1975)
  (c) T Joseph and N Sathyamurthy, *J. Indian Chem. Soc.* **62**, 874 (1985)
  (d) N Sathyamurthy, M Baer and T Joseph, *Chem. Phys.* **114**, 73 (1987)
  (e) J D Kress, R B Walker and E F Hayes, *J. Chem. Phys.* **93**, 8085 (1990)
  (f) J Z H Zhang, D L Yeager and W H Miller, *Chem. Phys. Lett.* **173**, 489 (1990)
  (g) B Lepetit and J M Launay, *J. Chem. Phys.* **95**, 5159 (1991)
  (h) N Balakrishnan and N Sathyamurthy, *Chem. Phys. Lett.* **201**, 294 (1993); **240**, 119 (1995)
  (i) K Sakimoto and K Onda, *Chem. Phys. Lett.* **226**, 227 (1994)

[3] A Kuppermann, J A Kaye and J P Dwyer, *Chem. Phys. Lett.* **74**, 257 (1980)
  J Manz, *Comm. At. Mol. Phys.* **17**, 91 (1985)

[4] V Balasubramanian, B K Mishra, A Bahel, S Kumar and N Sathyamurthy, *J. Chem. Phys.* **95**, 4160 (1991)

[5] D R McLaughlin and D L Thompson, *J. Chem. Phys.* **70**, 2748 (1979)
  T Joseph and N Sathyamurthy, *J. Chem. Phys.* **86**, 704 (1987)

[6] B B Mandelbrot, *The fractal geometry of nature* (Freeman, New York, 1982)

[7] J R Stine and R A Marcus, *Chem. Phys. Lett.* **29**, 575 (1974)
  N Agmon, *J. Chem. Phys.* **76**, 1309 (1982)

[8] A Lagana, M L Hernandez and J M Alvarino, *Chem. Phys. Lett.* **106**, 41 (1984)

[9] H R Mayne and R J Wolf, *Chem. Phys. Lett.* **81**, 508 (1981)

[10] A Rahaman and N Sathyamurthy, *J. Phys. Chem.* **98**, 12481(1994)

[11] S Mahapatra and N Sathyamurthy, *J. Chem. Phys.* **102**, 6057 (1995)

[12] S Mahapatra, R Ramaswamy and N Sathyamurthy, *J. Chem. Phys.* **104**, 3989 (1996)

[13] P Pechukas, *Chem. Phys. Lett.* **86**, 553 (1982)

[14] R D Levine and J L Kinsey, *Proc. Natl. Acad. Sci. (USA)* **88**, 11133 (1991)

[15] J Wilkie and P Brumer, *Phys. Rev. Lett.* **67**, 1185 (1991)

[16] Y Alhassid and R D Levine, *Phys. Rev.* **A46**, 4650 (1992)

[17] V A Mandelshtam, H S Taylor, C Jung, H F Bowen and D J Kouri, *J. Chem. Phys.* **102**, 7988 (1995)
  C C Marston, *J. Chem. Phys.* **103**, 8456 (1995)

[18] S Mahapatra and N Sathyamurthy, *J. Chem. Phys.* (submitted)
[19] R S Friedman and D G Truhlar, *Chem. Phys. Lett.* **183**, 539 (1991)
[20] M L Mehta, *Random matrices*, 2nd ed. (Academic, New York, 1990)
[21] S S M Wong and J B French, *Nucl. Phys.* **A198**, 188 (1972)
R Venkatraman, *J. Phys.* **B15**, 4293 (1982)
E Haller, H Köppel and L S Cederbaum, *Chem. Phys. Lett.* **101**, 215 (1983)
D M Leitner, R S Berry and R M Whitnell, *J. Chem. Phys.* **91**, 3470 (1989)
[22] S Mahapatra, *J. Chem. Phys.* **105**, 344 (1996)
[23] *Chaotic behaviour in quantum systems: theory and applications*, edited by G Casati (Plenum press, New York, 1985), Series B: *Physics* Vol. 120
O Bohigas, S Tomsovic and D Ullmo, *Phys. Rep.* **223**, 43 (1993)
Quantum chaos, edited by H Cerdeira, R Ramaswamy, G Casati and M C Gutzwiller (World Scientific Press, Singapore, 1991)
Quantum chaos, edited by G Casati and B V Chirikov, (Cambridge Univ. Press, Cambridge, 1994)
[24] M V Berry and M Tabor, *Proc. R. Soc. London* **A356**, 375 (1977)
[25] F J Dyson and M L Mehta, *J. Math. Phys.* **4**, 701 (1963)
[26] O Bohigas and M-J Gianonni, *Ann. Phys.* **89**, 393 (1975)
[27] T A Brody, J Flores, J B French, P A Mello, A Pandey and S S M Wong, *Rev. Mod. Phys.* **53**, 385 (1981)
Th Zimmermann, L S Cederbaum, H-D Meyer and H Köppel, *J. Phys. Chem.* **91**, 4446 (1987)
[28] O Bohigas, M-J Giannoni and C Schmidt, *Phys. Rev. Lett.* **52**, 1 (1984)
[29] M C Gutzwiller, *Chaos in classical and quantum mechanics* (Springer-Verlag, New York, 1991)
[30] R Ramaswamy, in *Schrödinger centenary surveys in physics*, edited by V Singh and S Lal (World Scientific,Singapore, 1988) pp 236
[31] A Pandey, *Ann. Phys.* (*NY*) **119**, 170 (1979)
[32] D M Leitner, H Köppel and L S Cederbaum, *J. Chem. Phys.* **104**, 434 (1996)
[33] Y Alhassid and N Whelan, *Phys. Rev. Lett.* **70**, 572 (1993)
[34] S Kumar, N Sathyamurthy and R Ramaswamy, *J. Chem. Phys.* **103**, 6021 (1995)
[35] See, *e.g.* R Ramaswamy, in *Reaction dynamics: recent advances*, edited by N Sathyamurthy (Narosa Press, New Delhi, 1990), pp 101–119
R Ramaswamy, in *Atomic and molecular physics*, edited by A P Pathak (Narosa Press, New Delhi,1992), pp 112–117
[36] K Someda, R Ramaswamy and H Nakamura, *J. Chem. Phys.* **98**, 1156 (1993)
[37] W H Miller, *Adv. Chem. Phys.* **25**, 69 (1974); **30**, 77 (1975)
[38] S R Channon and J L Lebowitz, *Ann. N. Y. Acad. Sci.* **357**, 108 (1980)
R S Mackay, J D Meiss and I C Percival, *Physica* **D13**, 55 (1984)
D Bensimon and L P Kadanoff, *Physica* **DL3**, 82 (1984)
M J Davis and S K Gray, *J. Chem. Phys.* **84**, 5389 (1986)
[39] F Gögtas, G G Balint-Kurti and A R Offer, *J. Chem. Phys.* **104**, 7927 (1996)
[40] A Carrington and R A Kennedy, *J. Chem. Phys.* **81**, 91 (1984)
A Carrington and I R McNab, *Acc. Chem. Res.* **22**, 218 (1989)
[41] M Berblinger, E Pollak and Ch Schlier, *J. Chem. Phys.* **88**, 5643 (1988)
E Pollak and Ch Schlier, *Acc. Chem. Res.* **22**, 223 (1989)
[42] T Turner, O Dutuit and Y T Lee, *J. Chem. Phys.* **81**, 347 (1984)
[43] J E Pollard, J A Syage, L K Johnson and R B Cohen, *J. Chem. Phys.* **94**, 8615 (1991)
J E Pollard, L K Johnson and R B Cohen, *J. Chem. Phys.* **95**, 4894 (1991)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 2,
February 1997
pp. 425–437

Chaos in intrinsic motion of nuclei and its role in dissipative collective dynamics

SANTANU PAL
Variable Energy Cyclotron Centre, 1/AF Bidhan Nagar, Calcutta 700 064, India

Abstract. We shall discuss the role of chaotic intrinsic motion in dissipative dynamics of the collective coordinates for nuclear systems. Using the formalism of linear response theory, it will be shown that the dissipation in adiabatic collective motion depends on the degree of chaos in the intrinsic dynamics of a system. This gives rise to a shape dependent dissipation rate for collective coordinates when the intrinsic motion is described by the independent particle model in a nucleus. The shape dependent chaos parameter measuring the degree of chaos in the intrinsic dynamics of the nuclear system will be obtained using the interpolating Brody distribution of nearest neighbour spacings in the single particle energy spectrum. A similar shape dependence is also found to be essential for phenomenological dissipation rates used in fission dynamics calculations.

Keywords. One-body dissipation; nuclear friction; chaos.

PACS Nos 24.60; 21.60; 05.45; 25.70

1. Introduction

The chaotic aspects of nuclear dynamics have invoked considerable interest [1–11] during the last decade. The nucleus is a particularly suitable candidate to study chaos because of its rich dynamics displaying both single-particle and collective features, interplay between microscopic and macroscopic degrees of freedom and the possibility of perturbing such a system with varying degrees of violence in a collision. It is worthwhile to clarify at this point that though the nucleus being a many body system of interacting fermions is expected to be fully chaotic, it is always not necessarily so. It may be recalled here that the fluctuation properties of low energy nuclear spectra follow those of eigenvalues of Gaussian Orthonormal Ensemble (GOE) of random matrices [12,13], which were subsequently found [14] to characterize the fluctuation statistics of eigenvalues of systems with nonintegrable classical motion. Fluctuation analyses of rotational spectra were performed [15] to gain insight into the structure of hot rotating nuclei. On the other hand, both chaotic and regular aspects were also found [16] in different regimes of nuclear spectra. Since many nuclear properties can be understood in terms of a relatively few degrees of freedom, the domain of the nuclear dynamics relevent for these properties can display chaotic features in varying degrees. In particular, the short range nature of the nucleon-nucleon interaction makes the mean field description of the nucleus a very good approximation. In its simplest form, independent particles moving in a potential well is known [17] to explain a number of nuclear bulk properties quite successfully. Thus in the mean field regime, the nuclear dynamics will be regular or chaotic depending on the nature of the mean field. While the single particle motion would be completely regular in a spherical potential, it would be more and more chaotic as the potential is progressively deformed [18], reaching the full chaotic limit for a strongly deformed nuclear shape. In the present article, we shall concern ourselves with the study of chaos in the nuclear single particle dynamics and its role in damping of nuclear collective motions.

Dissipation occurs in nuclear dynamics when the kinetic energy associated with the collective coordinates are irreversibly transferred to the intrinsic (non-collective) degrees of freedom of a nucleus. Such situations arise in the decay of the giant resonances, fission and in collisions between two heavy nuclei. Since collisions provide the opportunity to study the dissipation or damping phenomena under controlled conditions, the early motivation for developing theoretical models for dissipation in nuclei came from damped low energy heavy ion collisions [19]. Gross [20] pioneered the concept of one-body mechanism of dissipation in which the nucleons bouncing off the moving boundary of the time dependent average one-body potential soak energy from the collective motion resulting in a damping force in the collective degrees of freedom. Blocki and his collaborators [21] subsequently derived a simple expression namely the "wall formula" for one body dissipation. The wall formula was also obtained from a formal theory of one body nuclear dissipation based on classical linear response techniques [22]. The wall formula, in conjunction with the "window formula" [21] (to account for the role of nucleon exchange through the neck in a dinuclear system), was found to be quite successful in reproducing a large volume of experimental data of damped heavy ion collisions [23], fusion [24] and fission [21, 25].

One of the important assumptions of the wall formula concerns the randomization of the particle motion. It is assumed [21] that the successive collisions of a nucleon with the one body potential give rise to a velocity distribution which is completely random. In other words, a complete mixing in the classical phase space of the particle motion is required. This is normally satisfied for one body potentials whose shapes are rather irregular. It was early realised [21, 22] that any deviation from this randomization assumption would give rise to a reduced strength of the wall formula. This happens because the energy transferred to a particle from the time-dependent wall could be partly reversible if the motion of the particle is not completely random. In the present paper, we shall re-examine the classical wall formula in order to distinguish between the reversible and irreversible energy transfers. We shall identify the irreversible energy transfer as the true one-body dissipation. We shall further show that the original wall formula can be modified to describe the irreversible energy transfer. This modification would be obtained in the form of a scaling factor which would account for the degree of irregularity of the system. The modified wall formula thus obtained would be applicable to systems in which the particle motions are not fully randomized. In order to achieve the above, we shall exploit the characteristic features of the phase space of particle motion.

Recent progresses [26] in understanding the nature of dynamical systems have demonstrated that even simple systems possess a rich phase space structure and classical dynamics can be idealized for the two extreme cases of either fully regular (integrable) or fully chaotic (nonintegrable) motions. However, most of the systems of practical interest fall in between the above limiting cases and one has to often deal with "mixed" systems which display both regular and chaotic features. Arvieu *et al* [3] considered the classical motion of a particle moving in a deformed potential and pointed out the importance of the topology of phase space to characterise such motions. Heiss *et al* [27] studied single particle motion in axially deformed nuclei both classically and quantum mechanically. Considering shapes with quadrupole and higher multipole deformations, they succeeded in explaining the distinguishing features of shell structure in prolate and oblate shapes in terms of the underlying classical dynamics. In their detailed numerical study of classical particles in vibrating cavities of various shapes, Blocki *et al* [9] showed a strong correlation between chaos in classical phase space and the efficiency of energy transfer from the collective to the intrinsic motion. In particular, it was demonstrated in Ref. [9] that while the energy trasfer is much smaller than the wall formula limit in a cavity undergoing quadrupole vibration, it reaches the wall formula limit for higher multipole vibrations. Similar conclusions were also reached when the single particle motion was treated quantum mechanically [10]. As mentioned earlier, these essentially confirm the contention that the wall formula is valid when the intrinsic dynamics is fully chaotic.

In the present paper, we shall review some of our works [8, 18, 28, 29] aimed at a modification of the wall formula in order to make it amenable to mixed systems. It would be argued that while the energy transfer to the particle motion described by the regular part of the classical phase space is reversible, energy flows irreversibly to the particle motion which belongs to the chaotic phase space. We shall subsequently make use of a measure for the degree of chaos or nonintegrability for mixed systems which is defined as the relative volume of the phase space that belongs to the chaotic trajectories. We shall incorporate this measure of nonintegrability into the linear response formulation of the one body dissipation [22]. We would obtain a "scaled wall formula" which would be applicable for any system between a fully regular (no dissipation) and a fully chaotic one (original wall formula dissipation). Finally, we would present some numerical results to illustrate the effect of deformation on one body dissipation. We would further point out that a deformation dependent friction is found to be essential in fission dynamics calculations [30, 31] and the nature of this deformation dependence is very similar to our present theoretical prediction.

In the following sections, we shall first briefly review the classical linear response theory for one body dissipation, developed earlier by Koonin and Randrup [22], and then describe its modification [29] to obtain the scaled wall formula in § 3. In § 4, the scaling parameter or the measure of chaos will be numerically obtained for different deformed shapes. Finally, § 5 will contain a summary and discussions of our results.

2. One body dissipation revisited

We shall first briefly review the classical linear response theory of one body dissipation developed earlier by Koonin and Randrup [22]. The independent particle model of the nucleus will be considered here in which the motion of the individual nucleons in a time-dependent nuclear single particle potential is governed by the one body Hamiltonian $H(\mathbf{r}, \mathbf{p}; t)$. Here $\mathbf{r}$ and $\mathbf{p}$ denote the nucleon position and momentum respectively whereas t denotes the time. The Hamiltonian H may be split into a time independent part H_0 and a with $H_1(\mathbf{r},\mathbf{p};t=0)=0$.

Under the linear response approximations which require the validity of a perturbative treatment and the assumption that the relaxation time of the intrinsic motion is short in comparison with the time scale for collective motion, the rate of energy dissipation from the collective motion can be expressed as,

$$\dot{Q} = -\int \mathrm{d}\mathbf{r} \int \frac{\mathrm{d}\mathbf{p}}{(2\pi)^3} \left(\int_0^\infty \mathrm{d}t' \dot{H}_1(\mathbf{R}_0(\mathbf{r},\mathbf{p};t'),\mathbf{P}_0(\mathbf{r},\mathbf{p};t');t) \right) \dot{H}_1(\mathbf{r},\mathbf{p};t) \times \frac{\partial f_0(\mathbf{r},\mathbf{p})}{\partial H_0} \tag{2.2}$$

This equation corresponds to a physical picture in which a particle originating from a point $(\mathbf{r},\mathbf{p})$ in phase space contributes a dissipation rate equal to the product of the initial impulse received $\dot{H}_1(t)$ and the sum of all impulses received subsequently along its entire (unperturbed) trajectory $(\mathbf{R}_0,\mathbf{P}_0)$. In the above expression, f_0 is the single particle phase space distribution function governed by the unperturbed Hamiltonian H_0 and the factor $\partial f_0/\partial H_0$ ensures that for a Fermi-Dirac distribution only the particles near the Fermi surface contributes.

The assumption of a adiabatic collective motion required to arrive at Eq. (2.2) was critically examined in ref. [32] where the wall formula was obtained as the low-frequency limit of the one-body RPA damping. It was observed that the relaxation of the ideal adiabatic requirement to realistic collective speeds reduces the damping by 20%–30% Therefore the above eq. (2.2) shall give the leading contribution to one-body damping even when the collective and intrinsic time scales become comparable.

Considering a leptodermous system in which the nuclear potential is uniform throughout the volume but rises steeply at the surface, the time integral in eq. (2.2) can be written as a sum of the impulses received by a particle during its successive encounters with the nuclear surface along its unperturbed trajectory. Separating the contribution of the first impulse given to a particle near its point of origin at $t'=0$ (local part) from those arising out of the successive reflections from other regions of the nuclear surface (non-local part), Koonin and Randrup obtained the energy damping rate as,

$$\dot{Q} = \dot{Q}_{\text{local}} + \dot{Q}_{\text{non-local}} \tag{2.3}$$

where

$$\dot{Q}_{\text{local}} = \frac{\rho\bar{v}}{4\pi} \int u^2(a) \left(\int \mathrm{d}^2\Omega_a \right) \mathrm{d}^2 a \tag{2.4}$$

and

$$\dot{Q}_{\text{non-local}} = \frac{2\rho\bar{v}}{\pi} \int u(a) \left(\int \mathrm{d}^2\Omega_a \cos^2\theta_a \sum_b u(b)\cos\theta_b \right) \mathrm{d}^2 a \tag{2.5}$$

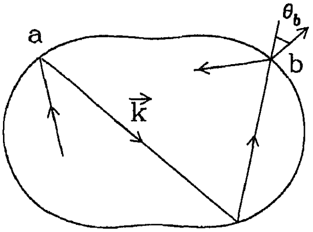


Figure 1. A typical trajectory in a deformed cavity.

In the above equations, we have followed the notations of ref. [22] where **k** is the nomentum of a particle as it leaves the surface point 'a' at an angle Ω_a, $u(a)$ is the ormal component of the surface velocity at 'a' and θ_b is the angle between the trajectory nd the normal at a point 'b' on the surface (figure 1). The summation is over all the ubsequent reflecting points 'b' along the trajectory after the first encounter at 'a'. ρ and $\bar{v}$ re respectively the nuclear mass density and the average nucleon speed inside the ucleus.

For regular systems such as a spherical cavity, it was shown in ref. [22] that the non-ocal contribution exactly cancels the local part. This essentially corresponds to an energy ransfer which is completely reversible. Thus the energy gained by a particle from the vall is eventually fed back to the wall when the particle motion is regular. The regularity n particle motion ensures complete reversibility under adiabatic conditions. An ntegrable system thus becomes completely non-dissipative in this picture. On the other and, the non-local part becomes zero for a system which is sufficiently irregular. This an also be seen from the eq. (2.5) where the surface velocity components at successive eflection points are completely random for a cavity of highly irregular shape and herefore cancel out in the summation. This in turn corresponds to a completely rreversible energy transfer. The dissipation rate is thus solely governed by the local term or an irregular system which now reduces to the wall formula,

$$\dot{Q}_{\text{wall}} = \rho\bar{v} \int u^2(a)\mathrm{d}^2a \tag{2.6}$$

It is to be noted further that the energy damping is also given by the wall formula rrespective of the system being regular or chaotic when the time available for damping s short compared to the interval between successive collisions of a particle with the vall. Such situations may arise when other competing processes reduce the interaction ime during which the one-body damping is important. If this interaction time is too short for any possible transfer of particle energy to the wall after the first collision, the net nergy transfer rate would be given by the local term or equivalently, the original wall ormula.

3. One body dissipation in mixed systems

We have considered in the above the one-body dissipation rates in completely regular or fully chaotic systems. However, most of the physical systems of interest do not belong to either of the two limits of full regularity or complete chaos. The dynamics of such systems display both the characteristic features of regularity (tori) and chaos (fog) in the classical phase space. To obtain the dissipation rate in such mixed systems, we shall proceed as follows.

We shall first distinguish between the regular and chaotic trajectories which contribute to the Eqs. (2.4) and (2.5) above. Originating from a given point on the cavity wall and moving in a given direction, a regular trajectory closes smoothly in phase space. On the other hand, another trajectory leaving the same point but in a different direction could be a chaotic one which does not close in the phase space. Let Ω_{1a} be the solid angle which contains all the chaotic trajectories originating from the surface point 'a' while Ω_{2a} is the corresponding solid angle for the regular trajectories ($\Omega_{1a} + \Omega_{2a} = 4\pi$). It may be noted here that each of these solid angles could be composed of disjoint components. The Eq. (2.3) then can be written as,

$$\begin{aligned}\dot{Q} = {} & \frac{\rho\bar{v}}{4\pi}\int u^2(a)\left(\int_{\Omega_{1a}} \mathrm{d}^2\Omega_a\right)\mathrm{d}^2a + \frac{2\rho\bar{v}}{\pi}\int u^2(a) \\ & \times \left(\int_{\Omega_{1a}} \mathrm{d}^2\Omega_a \cos^2\theta_a \sum_b u(b)\cos\theta_b\right)\mathrm{d}^2a + \frac{\rho\bar{v}}{4\pi}\int u^2(a)\left(\int_{\Omega_{2a}} \mathrm{d}^2\Omega_a\right)\mathrm{d}^2a \\ & + \frac{2\rho\bar{v}}{\pi}\int u^2(a)\left(\int_{\Omega_{2a}} \mathrm{d}^2\Omega_a \cos^2\theta_a \sum_b u(b)\cos\theta_b\right)\mathrm{d}^2a \end{aligned} \tag{3.1}$$

The last two terms on the right hand side of this equation represent the local and non-local contributions to the dissipation due to the trajectories which are regular. We shall assume at this point that the net dissipation due to the regular trajectories will be zero, similar to the spherical system whose all the trajectories are regular [22]. Though the non-dissipative nature of a regular system has been demonstrated only for a spherical system and a general proof for any regular system is not available to the best of our knowledge, the latter can be well anticipated on physical grounds and has been numerically confirmed [9] for several integrable systems. The zero dissipation should therefore be a generic property of the phase space of integrable systems and we expect it to hold for that part of the phase space of a mixed system which is similar to that of an integrable system.

We shall next consider the second term on the right side of the Eq. (3.1) which represents the non-local contribution of the chaotic trajectories. We immediately note that this term should also vanish due to the random nature of the surface velocity components at successive reflecting points as has been mentioned in the earlier section. We shall therefore be left with only the local term arising from the chaotic trajectories and obtain,

$$\dot{Q} = \rho\bar{v}\int \tilde{\mu}(a)u^2(a)\mathrm{d}^2a \tag{3.2}$$

where

$$\tilde{\mu}(a) = \frac{\Omega_{1a}}{4\pi}$$

and can be defined as the fraction of the trajectories leaving a surface point 'a' which are chaotic. Since we do not expect any correlation between $\tilde{\mu}(a)$ and $u^2(a)$, the above equation can be further reduced to,

$$\dot{Q} = \bar{\mu}\dot{Q}_{\text{wall}} \tag{3.3}$$

where

$$\bar{\mu} = \frac{\int \tilde{\mu}(a)\mathrm{d}^2 a}{S}$$

and S is the total surface area. Thus $\bar{\mu}$ is the average fraction of the trajectories which are chaotic when the sampling is done uniformly over the surface.

We have thus arrived at a scaled version of the wall formula where the dissipation rate depends on the degree of chaos in the single particle motion. For a completely irregular system, $\bar{\mu}$ is 1 and we recover the full wall formula whereas for a system with integrable single particle dynamics, there would be no chaotic trajectories ($\bar{\mu} = 0$) and hence no dissipation. For mixed systems, it will be necesary to calculate $\bar{\mu}$ separately to obtain the dissipation.

It is of interest to note here that a quantity (chaoticity/chaos parameter) defined in a similar fashion to $\bar{\mu}$ is often used in the literature [33, 34] to express the degree of irregularity in the dynamics of a system. This parameter, to be denoted henceforth as μ, is defined as the fraction of the phase space occupied by the chaotic trajectories. For a three-dimensional billiard, this essentially reduces to a uniform volume sampling to find the fraction of the trajectories which are chaotic. In what follows, we shall approximate $\bar{\mu}$ by μ to simplify our calculation. In order to justify this approximation, we first note that both the volume and surface samplings give the same limiting values of 0 and 1 respectively for fully regular or completely chaotic systems. Also, both the volume and surface averages would be equal for systems in which the chaotic trajectories are uniformly distributed. Moreover, as a regular cavity is deformed progressivly towards more irregular shapes, both μ and $\bar{\mu}$ also change monotonically from 0 to 1. We therefore do not anticipate any drastic difference between these two measures for mixed systems. However, it would still be of interest to make a precise comparison between μ and $\bar{\mu}$ which we plan to do in future.

4. Chaos parameter from energy spectrum

We have so far considered the chaos parameter μ (or $\bar{\mu}$) as a classically defined quantity which can be calculated by sampling trajectories in the classical phase space. However for the present purpose, we shall obtain the chaos parameter from a measure of the fluctuations of the single particle energy spectrum of the corresponding quantum system. This will be essentially done for two reasons. Firstly, the measure of chaos from quantum spectra is known [35] to be equivalent to the measure from classical phase space and secondly, the ready accessability of the quantum calculation to us. Thus the quantum calculation would give us a measure of chaos in the corresponding classical system. It may be noted that though a quantal formulation of dissipation has been obtained [36] earlier, the generic features of chaos in the quantum spectra has not been incorporated in such a formulation as yet.

For a regular or classically integrable system, it has been proved [37] that in the semiclassical limit, successive energy levels of the corresponding quantum hamiltonian arrive randomly. This results in a Poisson distribution $P(s)$ for the spacings (s) between the nearest neighbours in the energy spectra. For generic irregular systems with time-reversal symmetry, it has been conjectured [14] that spectral fluctuations are universally reproduced by the Gaussian Orthonormal Ensemble (GOE). For mixed systems, it has been argued [35] that in the semiclassical limit a spectrum should consist of regular and irregular parts that are associated with the classical regular and irregular regions in phase space. This was subsequently confirmed by Bohigas *et al* [38] in their detailed work on manifestation of classical phase structure in quantum systems. Assuming that irregular and regular regions yield energy level sequences with GOE and Poisson spacing distributions, respectively, and that the whole spectrum is generated by a statistically independent superposition of those sequences with respective weight factors μ and $(1-\mu)$, Berry and Robnik [39] obtained a semiclassical formula for $P(s)$ that smoothly interpolated between the Poisson $(\mu = 0)$ and GOE $(\mu = 1)$ distributions. One can thus extract a value of μ by a least-square fit of this parametrised $P(s)$ to the level-spacing distribution of a given spectrum. However, it has been subsequently pointed out [40] that the Berry–Robnik distribution converges extremely slowly to the semiclassical limit. It was further shown [40] that the phenomenological Brody distribution [41]

$$P_{\mathrm{Brody}}(\mu, s) = as^{\mu}\mathrm{e}^{-bs^{\mu+1}} \tag{4.1}$$

with $a = (\mu + 1)b$ and $b = \left[\Gamma(1 + \frac{1}{1+\mu})\right]^{1+\mu}$ is more appropriate in the near semiclassical regime.

Before we close this section, it is worthwhile to point out that several authors have discussed [42,43] other approaches to obtain smooth interpolation between Poisson and GOE distributions. French *et al* [42] considered increasing symmetry breaking in a given Hamiltonian as the driving mechanism to push the spacing distribution from Poisson to GOE. On the other hand, Lenz *et al* [43] proposed a random matrix ensemble displaying transition from Poisson to GOE spectra. However, an interpolation from Poisson to GOE based on classical phase space arguments is more appropriate for the present purpose. We shall therefore use Brody parametrisation in the present work.

4.1 *Chaos parameter for a two centre shell model*

We have earlier studied the spectral statistics of single particle motion in two centre shell model potentials [8,28]. The total Hamiltonian of a single particle in the combined field of two axially symmetric harmonic oscillators centred at $z_1 < 0$ and $z_2 > 0$ (the position of the barrier is at the origin) joined smoothly by a neck (figure 2) and with a spin-orbit potential is given as:

$$H = T + V + V_{SO} \tag{4.2}$$

where T is the kinetic energy, V denotes the sum of the oscillator and neck potentials and V_{SO} is the spin-orbit interaction. The details of the potentials are given in ref. [28]. The above potential is essentially characterised by the neck parameter $\epsilon = U_2/U_1$ and the separation between the two centres $R = z_2 - z_1$.

A large number of levels (≈ 150) were calculated in Ref. [28] and the best-fit Brody parameter was extracted over a range of values of the separation R. The system

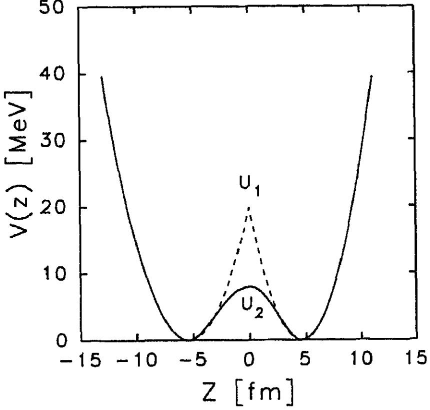


Figure 2. The two centre oscillator potential along the z-axis. The barrier heights without (dotted line) and with (full line) the neck potentials are U_1 and U_2, respectively.

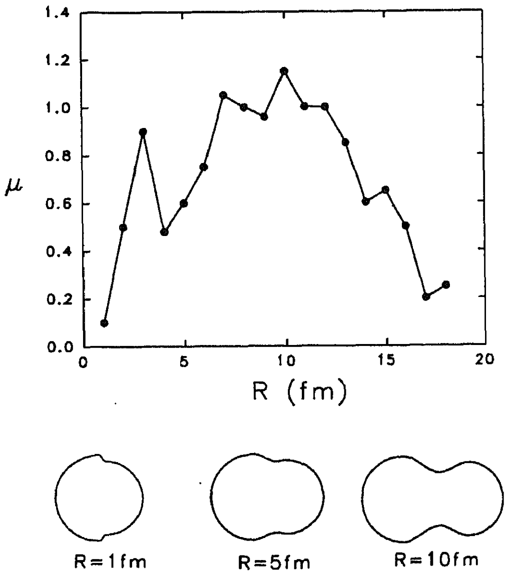


Figure 3. The variation of the chaos parameter μ with separation R for two centre shell potential. Dots (•) denote the calculated values and the lines are to guide the eye. Lower part shows the evolution of the profile of the equipotential surface. The neck parameter is held fixed at $\epsilon = 0.8$ for the cases above.

considered was of mass number $A = 260$ with an asymptotic mass asymmetry of 1.6. Figure 3 shows the variation of the chaos parameter with R. It is observed in this figure that the system is almost regular for near spherical shapes ($R \approx 0$) as expected. With increasing R, the single particle motion becomes increasingly chaotic as the system evolves towards a dinuclear complex. At still larger separations, the neck eventually snaps and two independent spherical nuclei emerge each supporting a regular single

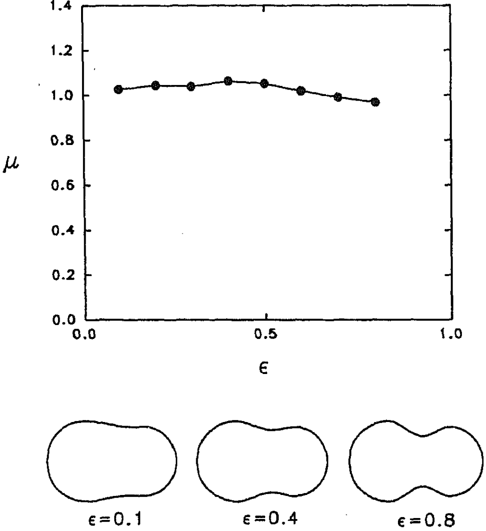


Figure 4. The variation of the chaos parameter μ with the neck parameter ϵ for two centre shell potential. Dots (•) denote the calculated values and the lines are to guide the eye. Lower part shows the evolution of the profile of the equipotential surface. The separation is held fixed at $R = 10\,\mathrm{fm}$ for the cases above.

particle motion. In fission or collisions between two heavy nuclei, the nuclear shape evolves in a manner similar to that depicted in figure 3. Therefore, the one-body dissipation that comes into play in the dynamics of such processes has to be scaled by the chaos parameter for the appropriate shapes. In particular, we note that the system is fully chaotic at $R \approx 10\,\mathrm{fm}$ which corresponds to the touching configuration of two nuclei of masses 100 and 160. Since in damped nuclear collisions, most of the damping occurs around this configuration, a scaling factor close to 1 is appropriate for the one-body dissipation in collisions. This, in turn, justifies the use of the original wall formula of Blocki *et al* in collision dynamics which, in fact, has been quite successful in reproducing [24] a large volume of experimental data.

On the other hand, energy dissipation takes place in fission when the nucleus evolves from a compact shape to the scission configuration. It has been observed from phenomenological studies [30, 31] of fission dynamics that a coordinate dependence of friction is essential for simultaneous reproduction of both the fission probability and the pre-scission neutron multiplicity. A reduced friction at compact shapes followed by a gradual increase upto the scission point was used successfully in the above studies. This shape dependence is exactly similar to our findings as discussed earlier. Therefore, the shape dependence of nuclear friction arising out of the chaoticity of the single particle states offers a possible explanation of the phenomenological [30, 31] shape dependence.

After studying the separation R dependence of the single particle chaos, it is of interest to study how it depends on other shape parameters. In particular, we would be interested to find the dependence of the chaos parameter on the neck parameter ϵ since the negative curvature in the potential profile in the neck region provides a defocussing effect on the single particle motion. With this view, the chaos parameter is obtained for a number of

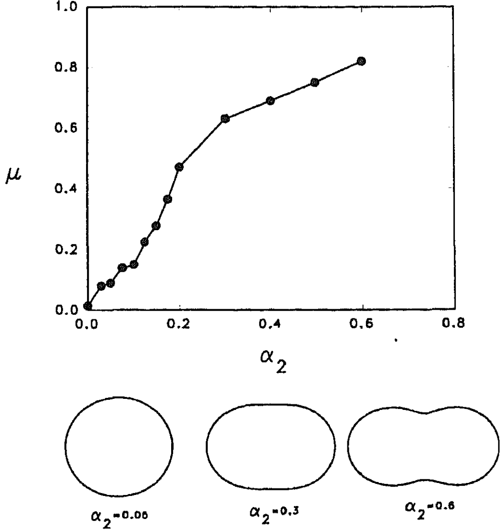


Figure 5. The variation of the chaos parameter μ with deformation α_2 for quadrupole cavities. Dots (●) denote the calculated values and the lines are to guide the eye. Lower part shows the evolution of the cavity shape with increasing deformation.

values of the neck parameter (for a given separation) and the results are given in figure 4. It shows that the single particle motion is fully chaotic over a broad range of the neck potential. This implies that the choice of the neck parameter is not very critical so far as its role in inducing chaos is concerned.

4.2 *Chaos parameter for a deformed cavity*

We shall now consider single particle motion in an axially symmetric cavity with quadrupole deformation. The main motivation for considering this system is twofold. Firstly, we would like to study systems for which larger number of levels can be numerically obtained and secondly, the relevence of quadrupole deformed shapes for the study of giant resonances in nuclei [29]. Specifically, we shall consider a cavity whose surface is deformed according to the Legendre polynomial of second order,

$$R(\theta) = \frac{R_0}{\lambda}[1 + \alpha_2 P_2(\cos\theta)] \tag{4.3}$$

where R_0 is the radius of a sphere having the same volume as the deformed cavity, λ is a volume preserving factor and α_2 is the strength of the quadrupole deformation. The single particle eigenenergies were obtained in ref. [18] by numerically solving the Schrodinger equation. The nearest neighbour level-spacing distributions were obtained from these eigenenergy sequences. The spacing distributions were subsequently fitted by the Brody distribution to yield the chaos parameter μ for cavities of various deformations.

We shall presently extend the above study to a wider range of deformation. A better description of the variation of the chaos parameter with deformation will be obtained here compared to the two-centre potential system. This is essentially because the chaos parameter is obtained here at smaller intervals of the deformation and larger number of levels (≈ 1000) are employed compared to the two-centre potential calculation. The variation of the chaos parameter with deformation is shown in figure 5. The nature of the variation is similar to the two-centre potential system, namely a regular to chaos transition is observed as the system is progressively deformed. It is of interest to note here that in typical giant quadrupole oscillations of nuclei, the amplitude of the deformation lies within 0.1 and one finds in figure 5 that the chaos parameter is small ($\mu \leq 0.3$) in this range. The one-body mechanism for the damping of giant resonances is, therefore, a very weak mechanism [29].

5. Summary and discussion

In the preceding sections, we have considered the wall formula for one body dissipation by relaxing the 'full randomization' assumption in order to make it applicable to systems in which the mixing in phase space is only partial. In particular, we investigated the effect of irregularity in the shape of the one body potential on the dissipation mechanism. Closely following the theoretical framework of one body dissipation developed earlier by Koonin and Randrup [22], we arrived at a dissipation rate which is a scaled version of the wall formula. We could show that the scaling factor is determined by a measure of chaos in the single particle motion.

Subsequently we calculated this measure for a two-centre potential system and a quadrupole deformed cavity. In both the cases, an order to chaos transition in the single particle motion was observed with increasing deformation of the system. This also corresponds to a gradual increase of the one-body dissipation as the system is progressivly deformed. Since the one-body dissipation is expected to be the dominant mechanism in slow collective processes such as low energy nucleus-nucleus collisions and fission, it is of interest to note that a shape dependent friction is also found to be essential for simultaneous reproduction of both the fission probabilty and the pre-scission neutron multiplicity in fission dynamics calculations [30, 31]. It was found in these works that a phenomenological friction which is very weak for compact configurations and increases after the neck is formed in a fissioning nucleus can describe a large number of experimental data. Interestingly, this corresponds exactly to our own conclusions regarding the shape dependence of nuclear friction. It is thus quite likely that the original wall formula determined from a series of simplifying assumptions may require some readjustments to account for the relaxtion of these assumptions towards more realistic situations. The present article describes some of the efforts in this direction and it is expected that the effects of shape-dependence would serve as an essential input in evolving a better model of one-body dissipation.

Acknowledgements

It is a matter of great pleasure to acknowledge collaboration with A K Chaudhuri, Vaishali Banerjee, Debabrata Biswas and T Mukhopadhyay during different phases of the work reported here.

References

[1] A Abul-Magd and H A Weidenmuller, *Phys. Lett.* **B162**, 223 (1985)
[2] J F Shriner Jr., G E Mitchell and E G Bilpuch, *Phys. Rev. Lett.* **59**, 435 (1987)
[3] R Arvieu, F Brut, J Carbonell and J Touchard, *Phys. Rev.* **A35**, 2389 (1987)
[4] O Bohigas and H A Weidenmuller, *Ann. Rev. Nucl. Sci.* **38**, 421 (1988)
[5] Y Alhassid, A Novoselsky and N Whelan, *Phys. Rev. Lett.* **65**, 2971 (1990)
[6] W J Swiatecki, *Nucl. Phys.* **A488**, 375 (1988)
[7] H A Weidenmuller, *Nucl. Phys.* **A520**, 509c (1990)
[8] S Pal and A K Chaudhuri, *Nucl. Phys.* **A537**, 237 (1992)
[9] J Blocki, J-J Shi and W J Swiatecki, *Nucl. Phys.* **A554**, 387 (1993)
[10] J Blocki, J Skalski and W J Swiatecki, *Nucl. Phys.* **A594**, 137 (1995)
[11] E Caurier, J M G Gomez, V R Manfredi and L Salasnich, *Phys. Lett.* **B365**, 7 (1996)
[12] A Bohr and B R Mottelson, *Nuclear structure*, (Benjamin, New York, 1969) Vol. I, Appendix 2C
[13] T A Brody, J Flores, J B French, P A Mello, A Pandey and S S M Wong, *Rev. Mod. Phys.* **53**, 385 (1981)
[14] O Bohigas, M J Giannoni and C Schmit, *Phys. Rev. Lett.* **52**, 1 (1984)
[15] T Dossing, B Herskind, S Leoni, A Bracco, R A Broglia, M Matsuo and E Vigezzi, *Phys. Rep.* **268**, 1 (1996)
[16] M T Lopez-Arias, V R Manfredi and L Salasnich, *Nuovo Cimento* **5**, (1994)
[17] A Bohr and B R Mottelson, *Nuclear structure*, (Benjamin, New York, 1969) Vol. I, 137
[18] T Mukhopadhyay and S Pal, *Nucl. Phys.* **A592**, 291 (1995) and *Nucl. Phys.* **A596**, 713 (1996)
[19] W U Schroder and J R Huizenga, *Treatise on heavy-ion science*, edited by D A Bromley (Plenum Press, 1984) Vol. 2, p. 115
[20] D H E Gross, *Nucl. Phys.* **A240**, 472 (1975)
[21] J Blocki, Y Boneh, J R Nix, J Randrup, M Robel, A J Sierk and W J Swiatecki, *Ann. Phys. (N.Y.)* **113**, 330 (1978)
[22] S E Koonin and J Randrup, *Nucl. Phys.* **A289**, 475 (1977)
[23] J Blocki, M Dworzecka, F Beck and H Feldmeier, *Phys. Lett.* **B99**, 13 (1981)
[24] J R Birkelund, L E Tubbs, J R Huizenga, J N De and D Sperber, *Phys. Rep.* **56**, 107 (1979)
[25] A J Sierk and J R Nix, *Phys. Rev.* **C21**, 982 (1980)
[26] Martin C Gutzwiller, *Chaos in classical and quantum mechanics*, Springer (Berlin), 1990
[27] W D Heiss, R G Nazmitdinov and S Radu, *Phys. Rev. Lett.* **72**, 2351 (1994); *Phys. Rev.* **C52**, 1179 (1995); *Phys. Rev.* **C52**, 3032 (1995)
[28] D Biswas, S Pal and A K Chaudhuri, *Phys. Rev.* **A46**, 6817 (1992)
[29] S Pal and T Mukhopadhyay, *Phys. Rev.* **C54**, 1333 (1996)
[30] P Froebrich, I I Gontchar and N D Mavlitov, *Nucl. Phys.* **A556**, 281 (1993)
[31] P Froebrich and I I Gontchar, *Nucl. Phys.* **A563**, 326 (1993)
[32] C Yannouleas, *Nucl. Phys.* **A439**, 336 (1985)
[33] J Blocki, F Brut, T Srokowski and W J Swiatecki, *Nucl. Phys.* **A545**, 511c (1992)
[34] Th Zimmermann, H-D Meyer, H Koppel and L S Cederbaum, *Phys. Rev.* **A33**, 4334 (1986)
[35] I C Percival, *J. Phys.* **B6**, L229 (1973)
[36] S E Koonin, R L Hatch and J Randrup, *Nucl. Phys.* **A283**, 87 (1977)
[37] M V Berry and M Tabor, *Proc. R. Soc. London* **A356**, 375 (1977)
[38] O Bohigas, S Tomsovic and D Ullmo, *Phys. Rep.* **23**, 43 (1993)
[39] M V Berry and M J Robnik, *J. Phys.* **A17**, 2414 (1984)
[40] T Prosen and M B Robnik, *J.Phys.* **A26**, 2371 (1993)
[41] T A Brody, *Lett. Nuovo Cimento* **7**, 482 (1973)
[42] J B French, V K B Kota, A Pandey and S Tomsovic, *Ann. Phys. (N.Y.)* **181**, 198 (1988)
[43] G Lenz and F Haake, *Phys. Rev. Lett.* **67**, 1 (1991)
G Lenz, K Zyczkowski and D Saher, *Phys. Rev.* **A44**, 8043 (1991)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 2,
February 1997
pp. 439-457

Chaos and localization in coupled quartic oscillators

M S SANTHANAM, V B SHEOREY, A LAKSHMINARAYAN
Physical Research Laboratory, Navarangapura, Ahmedabad 380 009, India

Abstract. We discuss some of the models for eigenfunction localization in Hamiltonian systems. In particular, we review some of our work on classical parametric scaling of orbits and identification of localized states in a two-dimensional quartic oscillator system which is deep in the classically chaotic region. We show that visual methods are a necessary complement to quantitative methods based on information entropies. Our preliminary results indicate that the periodic orbit stability determines the degree of localization in a class of states, even when the stable regions are of negligible measure.

Keywords. Quantum chaos; wave functions; localizations.

PACS Nos 05.45; 03.65

1. Introduction

The study of quantization of chaotic systems has been an active field of research for very many years now, and has seen substantial developments in the theoretical and experimental techniques. The celebrated trace formula of Gutzwiller [1] connects the quantum spectra semiclassically to unstable periodic orbits embedded in a sea of chaos. Random matrix theories (RMT) have been found to apply to remarkably low-dimensional but chaotic systems, while their putative range was many body physics. Experiments have been done in many areas from atomic, molecular systems to mesoscopic systems for which the concepts of 'quantum chaos' are essential. As is often pointed out, while we have a reasonable grip over the eigenvalues of a quantum spectra, via the Gutzwiller formula, eigenfunctions are still a region of mystery. There have been many computations of eigenfunctions, and even some experimental realizations have been made, but their real nature remains obscure.

We deal mainly with eigenfunctions and seek to understand the most remarkable of them, which we believe will also have the largest experimental impact. Towards this end we study smooth Hamiltonian systems in two degrees of freedom, and deal with homogeneous potentials. We are able to compute eigenfunctions which are in the highly excited regions of the spectrum and for parameter ranges which take us deep into classical chaos. We note that in contrast to previous work the eigenfunctions computed really probe the semiclassical regime. We have noticed the predominance of states which show density enhancements near unstable periodic orbits called 'scars', but also have studied the random matrix properties of these states, and we find that paradoxically even some strongly scarred states can show random matrix like distributions. There are however embedded in this 'chaotic' or irregular spectrum a localized.

Percival's early conjecture [2] that there are two types of states in a typical spectrum, the regular and the irregular, has been amply verified by later computations. However we note that the series we study here is not associated with a classically regular region of phase space although they have all the hallmarks of regular states. We associate simple periodic orbits to these states and study the effect of classical stability oscillations in the regime of *extreme* chaos. We focus attention in this paper on the study of these eigenfunctions, and since they are strongly localized, we wish to place them in the context of the general localization lore.

One of the distinguishing features of quantum mechanics is the non-spreading of the eigenfunctions in the whole of the classically accessible configuration or momentum space, succinctly called localization. Essentially, this is the tendency of the quantum eigenfunctions to accumulate only on certain subregions of the classically allowed region. In recent years, many studies on eigenfunctions, particularly of classically chaotic systems, have focussed on this particular aspect of quantum mechanics [1, 5, 11]. Quantum localization is a generic phenomenon in many areas of physics. In condensed matter physics, Anderson's model [3] is one of the earliest attempts to explain localization for the motion of an electron in a random alloy, in which potential well depths at each lattice site are random. In this case it has been shown that the wavefunctions in configuration space are exponentially localized. In quantum optics, Kuklinski [4] has shown that Anderson type localization also occurs in a photon number space if sharp pulses of coherent light interact with an optical cavity with non-linear susceptibility. It has been suggested that this localization can be used to generate an electromagnetic field in a pure Fock state. Atomic and molecular systems provide many instances of localization phenomena. Using Husimi distributions Müller and Wintgen [6], have studied the localization in the case of the hydrogen atom in a uniform magnetic field and connected most of the eigenstates to the classical periodic orbits and invariant manifolds. Microwave ionization of hydrogen atom from its highly excited states has been the subject of active investigation, both theoretically and experimentally. Theoretical studies on an one-dimensional model predicted [7] that for hydrogen atoms in microwave fields, as a result of localization, there would be an increase in quantum ionization threshold over and above the classical threshold, beyond a critical value of the microwave field. This has been experimentally verified [8] . More recently, Wilkinson *et al* [9] using tunnel-current spectroscopy in semiconductor heterostructures were able to experimentally observe the effect of localized states.

2. Localization in chaotic quantum systems

Localization has been used to denote phenomena in a wide class of problems, and there is probably no consensus on what exactly this term means. Even in the context of quantization of chaotic systems, localization could mean the scarring of eigenfunctions due to simple orbits, or something even more stringent. There is perturbative localization [10], dynamical localization [11], Anderson localization [3], scarring localization [33], and a bewildering array of mechanisms to go with them. We will focus on localization which is close to scarring localization, however, as will be apparent this is a strong form of scarring localization, that one may attribute also to adiabatic mechanisms, and is exponential in some sense.

Below we study localization in model quantum systems whose classical limit is chaotic. Study of localization is important since it is a pure quantum phenomenon whose mechanism and systematics are yet to be understood in complete detail. Secondly, from the point of view of quantum chaos, studies on deviations of theoretical predictions for energy level statistics in the presence of localized states have been done [12].

Based on the semiclassical eigenfunction hypothesis and Shnirelmann's theorem [13], Berry [14] and Voros [15] argued that a typical eigenfunction in the irregular quantum regime is ergodic over the entire configuration space. At first sight, it would appear that since the underlying classical dynamics of the chaotic systems is ergodic, any wavepacket launched on any of the infinite periodic orbits will spread over all the accessible space and hence would not support localized eigenfunction structures. First results contrary to the above hypothesis came from the early work of McDonald and Kaufmann [16], from the studies on eigenfunction structures of stadium billiards. Heller [17], using semiclassical wave propagation techniques, argued that scars can occur near the *least* unstable periodic orbits. In chaotic systems, all periodic orbits are unstable and they contribute significantly to density enhancements in the vicinity of the orbit [18]. In this study it was also shown that wavefunctions averaged over small energy intervals can be related to a sum over periodic trajectories in the semiclassical limit.

Other model systems, like nonlinearly coupled quartic oscillators, kicked rotor and an atom in a magnetic field were studied and they also showed sharp localization. The kicked rotor has been extensively investigated both classically and quantum mechanically [19, 33]. The Hamiltonian of the periodically kicked rotor is given by

$$H(t) = \frac{p^2}{2I} + \frac{\mu}{T} IV(\theta) \sum_{n=-\infty}^{\infty} \delta(t - nT), \tag{1}$$

where $V(\theta) = \cos(\theta)$, I is the moment of inertia, T is the kicking period and classically μ is the only effective parameter in the model and it represents the kick strength. Classically, for large values of kick strength, i.e $\mu > 1.0$ the dynamics is chaotic. The quantum mechanical problem is solved by constructing the Floquet operator and solving the resulting eigenvalue problem in the momentum representation. After some mathematical transformations, the Schrödinger equation for the kicked rotor in momentum representation looks similar to the one for Anderson model for the motion of an electron in a crystalline lattice of random potential wells. Thus, by mapping the kicked rotor problem to the Anderson's problem, Grempel *et al* [20] attributed localization in the kicked rotor to destructive quantum phase interference since the same mechanism is responsible for localization in Anderson's model. However, for all their avowed similarity, Anderson's model and the kicked rotor are not the same. The potential in kicked rotor is not strictly random, it is at best pseudo random and the quasi-energy of the rotor becomes a parameter of the potential.

Rydberg atoms in strong magnetic fields have emerged as one of the testing grounds for quantum chaos. The interaction for hydrogenic atom in strong magnetic field contains a quadratic Zeeman term which makes the problem non-trivial. We consider a particular magnetic quantum number manifold, say $m = 0$, and assume the magnetic field $\mathbf{B}$ to be directed along the z axis. In such a case, the Hamiltonian written in cylindrical coordinates is,

$$H = \frac{1}{2M}(p_\rho^2 + p_z^2) - \frac{e^2}{(\rho^2 + z^2)^{1/2}} + \frac{1}{8}M\omega_c^2\rho^2, \tag{2}$$

where ϕ motion has been separated and $\omega_c = eB/Mc$ is the cyclotron frequency. Rewriting it in semi-parabolic coordinates [21] we get,

$$H = \tfrac{1}{2}(p_u^2 + p_v^2) - 4\epsilon(u^2 + v^2) + 8u^2v^2(u^2 + v^2), \tag{3}$$

where the scaled energy $\epsilon = E\gamma^{-2/3}$ is the parameter and $\gamma = B/B_0$ with $B_0 = 2.35\times 10^5$ T. It is to be noted that in this form this system resembles the Hamiltonian for a particle in a two-dimensional non-linear potential with ϵ as the parameter. Many authors [21] have attempted this problem theoretically as well as experimentally. Rau [10] has suggested that localization in this problem can arise from the perturbative breaking of degeneracy. The integrable part of the atomic many body system, at high excitation, displays a very high degree of degeneracy. For example, in the integrable quantum hydrogen atom problem, states with the same principal quantum number n, but with different angular momentum quantum numbers l, are degenerate and form a manifold or a subspace in the full Hilbert space. Perturbations to these integrable systems lifts these degeneracies and is accompanied by mixing of base states of these manifolds. Since angular momentum states mix and diffuse, the uncertainty principle restricts wavefunctions to be localized spatially because the angle and angular momentum are conjugate variables. Furthermore it has been shown that this perturbative localization is common to the asymmetric rotor and to a wide class of few body problems in atomic and nuclear physics.

3. The coupled quartic oscillator

3.1 *Classical aspects*

One of the most widely studied models of quantum chaos is the non-linearly coupled quartic oscillator (henceforth referred to as quartic oscillator) [11, 22, 43]. We study the Hamiltonian given by

$$H(p_x, p_y, x, y; \alpha) = p_x^2 + p_y^2 + x^4 + y^4 + \alpha x^2y^2, \tag{4}$$

where α is a parameter. This Hamiltonian is of interest in $SU(2)$ Yang-Mills field theory in the limiting case when $\alpha = \infty$. Since the system is homogeneous the classical motion scales with energy. This implies that from the knowledge of classical dynamics at one energy, dynamics at any other energy can be obtained [23]. Though the classical dynamics of this system is reported extensively [43, 29], we recollect some of its salient features. The potential is bounded for all values of α greater than -2, except at $\alpha = \infty$. In this limit when $\alpha = \infty$, though classically the potential is open, the quantum energy spectrum is discrete. The system is integrable as well as separable for $\alpha = 0, 2, 6$. The Poincaré section shows predominance of irregular trajectories as α increases beyond 6. Though the regular region shrinks with an increase in α, due to bounded stability oscillations of the channel periodic orbit, there are certain windows of α values for which an infinitesimally small regular region around the orbit exists, even for very high values of α. The duality in the parameter space enables us to map the interval [0, 2] onto [6, 2] and the interval [−2, 0] onto [∞, 6] and

$$\frac{H(p_x,p_y,x,y,;\alpha)}{(\alpha+2)^{1/6}} = \frac{H(p_u,p_v,u,v,;\alpha')}{(\alpha'+2)^{1/6}}, \tag{5}$$

where

$$\alpha' = \frac{(12-2\alpha)}{(2+\alpha)}. \tag{6}$$

Geometrically, this transformation given by

$$u = (4+2\alpha)^{1/6}\,\frac{(x+y)}{2}, \quad v = (4+2\alpha)^{1/6}\,\frac{(x-y)}{2}, \tag{7}$$

corresponds to rotation of the potential by $\pi/4$ and a scaling.

We have recently shown that the Hamiltonian in (4) also *locally* scales in the vicinity of the channel periodic orbit *with respect to the parameter* α [26]. The stability of the simple channel periodic orbit $(x, y=0, p_x, p_y=0)$, as measured by $\mathrm{Tr}\,J(\alpha)$, where $J(\alpha)$ is the monodromy matrix, displays bounded oscillations as a function of α. The formula due to Yoshida [24] for the trace of the monodromy matrix, applied to the case of the quartic oscillator in (4) gives

$$\mathrm{Tr}\,J(\alpha) = 2\sqrt{2}\cos\left(\frac{\pi}{4}\sqrt{1+4\alpha}\right). \tag{8}$$

The orbits are stable if $|\mathrm{Tr}\,J(\alpha)| < 2$ and are unstable if $|\mathrm{Tr}\,J(\alpha)| > 2$. By inverting this formula using these stability conditions we find that at values of $\alpha = n(n+1)$, where n is an integer, the channel orbit changes stability through a pitchfork bifurcation. Figure 1 shows the stability curves for the channel orbit and the 45^0 diagonal orbit. The channel orbit corresponds to positive α range and the diagonal orbit to negative α values, as per the transformation in (6). It is useful to note that the Hamiltonian bifurcation theory applied to two-dimensional conservative systems, allows for five and only five types of bifurcations [25].

We have numerically demonstrated that if the stability at two different values α and α' are equal, i.e $\mathrm{Tr}\,J(\alpha) = \mathrm{Tr}\,J(\alpha')$, then the Poincaré section locally scales around the vicinity of the channel orbit. If the system possesses reflection symmetries about a particular axes, say the x axis, then the Poincaré section simplifies to the condition $x = 0$ with either sign of p_x. This constitutes a half-Poincaré map [25] and has all the properties of the conventional Poincaré map. Let

$$q_2' = f(q_2,p_2;\alpha), \quad p_2' = g(q_2,p_2;\alpha) \tag{9}$$

be the half-Poincaré map, with $\alpha' > \alpha$. Due to the reflection symmetries of the oscillator, $f(q_2,p_2;\alpha) = -f(-q_2,-p_2;\alpha)$, and similarly for g. Since the Poincaré section scales, the

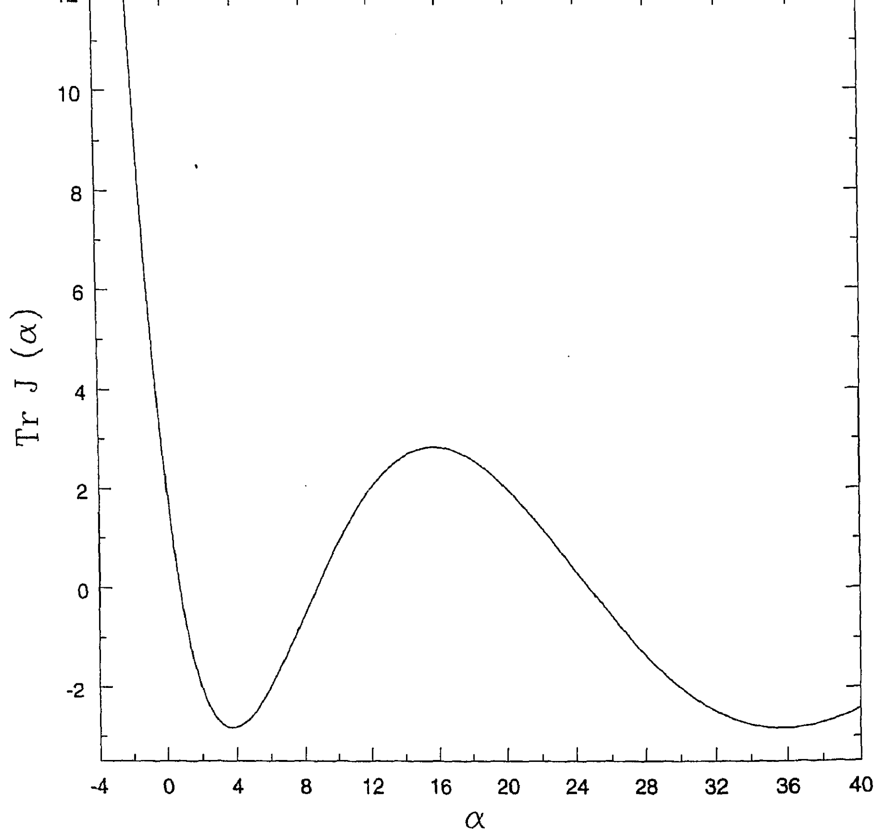


Figure 1. The stability oscillations of the channel orbit for the Hamiltonian (4) as a function of α.

above functions also will scale as follows,

$$\left(\frac{\alpha'}{\alpha}\right)^{-\gamma_1} f(q_2,p_2;\alpha) = f\left(\left(\frac{\alpha'}{\alpha}\right)^{-\gamma_1} q_2, \left(\frac{\alpha'}{\alpha}\right)^{-\gamma_2} p_2;\alpha'\right), \tag{10}$$

where γ_1 and γ_2 are scaling exponents with values close to 0.625 and 0.325 respectively. In fact, we have presented evidence that is more general and covers all the two-dimensional homogeneous Hamiltonian systems whose Poincaré sections around channel orbits scale with α [26]. We have demonstrated that for the class of Hamiltonians,

$$H_{2n} = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 + \frac{1}{2n}(\beta_1 q_1^{2n} + \beta_2 q_2^{2n}) + \frac{\alpha}{2}(q_1^2 q_2^{2n-2} + q_2^2 q_1^{2n-2}), \tag{11}$$

the scaling exponents depend only on the homogeneity of the potential as follows,

$$\gamma_1 = \frac{2n+1}{4n}, \quad \gamma_2 = \frac{2n-1}{4n}. \tag{12}$$

For another class of Hamiltonians of the form,

$$H'_{2n} = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 + \frac{1}{2n}(\beta_1 q_1^{2n} + \beta_2 q_2^{2n}) + \frac{1}{n}\alpha q_1^n q_2^n, \quad n > 2, \tag{13}$$

the channel orbit is always marginally stable independent of α. The stable region around this orbit scales with α and the scaling exponents are given by

$$\gamma_1 = \gamma_2 = \frac{1}{n-2}. \tag{14}$$

In §7 we will see that stability oscillations of the channel periodic orbit has implications for wavefunction localization.

3.2 *Quantum aspects*

Quantum mechanically we solve the Schrödinger equation,

$$H(x, y; \alpha)\Psi_n^\alpha = E_n^\alpha \Psi_n^\alpha \tag{15}$$

using symmetrized linear combination of one-dimensional quartic oscillator eigenfunctions given by

$$\Psi_j(x, y) = \mathcal{N}(n, m)\{\phi_n(x)\phi_m(y) + \phi_m(x)\phi_n(y)\}, \tag{16}$$

where $\mathcal{N}$ is the normalization constant depending on n and m, both even integers and the index j represents a unique pair (n, m) of these integers. The functions $\phi_n(x)$ are obtained numerically by accurately solving a one-dimensional Schrödinger equation for bound states of quartic oscillator (see [27] and references therein). Using 160 to 175 one-dimensional quartic oscillator basis states, we numerically diagonalized a large Hamiltonian matrix of orders ranging from about 12000 to 15000 on a IBM RISC 6000/580 machine to obtain about 2000 converged eigenvalues and eigenfunctions. In all our calculations, positive values of α in the range $[6, \infty]$ were used since more eigenvalues converged in this case than for the corresponding range $[0, -2]$. A detailed account of numerical techniques for large matrix diagonalization and its convergence problems will be presented elsewhere.

This potential belongs to the C_{4v} symmetry group, namely the group of all the transformations of a square. This group has four one-dimensional irreducible representations, but for all our calculations reported in this paper we confined ourselves to A_1 representation, which is symmetric under reflections on the axes and diagonals. In fact, the basis function given in (16) transforms according to this representation. By choosing such appropriately symmetrized basis functions the Hamiltonian matrix is block diagonal.

Since the system in (4) classically scales, the quantum spectra at different values of Planck's constant are related by

$$E(\hbar; \alpha) = C^4 E(\hbar/C^3; \alpha), \tag{17}$$

where C is a constant. The duality in parameter space gives us one more relationship between quantum energy levels between different values of α related by (6):

$$\frac{E(\hbar; \alpha)}{(2+\alpha)^{1/6}} = \frac{E(\hbar; \alpha')}{(2+\alpha')^{1/6}}. \tag{18}$$

4. Localized states in quartic oscillator

In the integrable quartic oscillator all the eigenfunctions are localized on the invariant tori. Here we address the question of systematics and morphology of highly excited states when the classical system is largely chaotic. Structurally, most of the localized states of the quartic oscillator display one or more of the following three kinds of features, namely, (i) localization along the channel periodic orbits, (ii) localization along the orbit 45° to the channel states, (iii) localization along a circle, called ring states. However, there are other scarred states with interesting structures which do not fall in any of the above categories. At the moment, many of these scarred states can only be visually identified since there are no quantitative measures to identify them. One of the straightforward methods of identifying a scarred state is to look at it visually. To this end, we define a configuration space density

$$\rho_n^\alpha(x,y) = |\Psi_n^\alpha(x,y)|^2. \tag{19}$$

By plotting $|\Psi(x,y)|^2$ as an intensity plot, we identify the scarred states. In one of the first extensive studies on the eigenfunctions of the quartic oscillator, Eckhardt *et al* [28] have reported eigenfunctions in configuration space of sixty states starting from the ground state, which shows all the familiar localized structures mentioned above. They have identified the classically unstable periodic orbits that produce these localized structures. Probing much deeper in the semiclassical regime, even as high as up to 2000 states that we computed for various α values from 60 to 96, we found that the localized states scarred by the channel orbits continue to exist. These special states scarred by channel orbits form a dominant series in the semiclassical region, which implies that it is most easily amenable to semiclassical treatment as will be seen later. We showed the existence of another class of scarred states called *shadow states*, which are weakly scarred by the channel orbit in addition to some other nearby orbits and hence show weak localization features in comparison with its dominant counterpart. These states occur near the channel localized states and form another series in the spectrum which seems to be absent at the lower end of the spectrum. Another peculiar set of states were those which localized on a point and consequently intensity in the rest of the configuration space was weak. Some intensity plots of channel localized and shadow states and of four weak states have been published [29].

There are two important points to be noted. Firstly, many Hamiltonian systems in the form of (11) allow for these types of localized states. For instance, the potential given by $V(x,y) = x^2 + y^2 + \beta x^2 y^2$, where β is a parameter, also shows these familiar localized structures [30]. Secondly since the quantitative measures to identify the scarred states are lacking, the visualization of eigenfunctions, at the moment, seems to be the viable alternative. In fact, as we will see later, the visual techniques complement the numerical methods, even in such cases where certain measures do identify particular kind of scarred states.

5. Quantitative methods for localized states

The states scarred by channel orbits form a small fraction of the total number of eigenstates we computed. There are about 60–70 channel localized states in the first 2000 states, many others exhibit various other types of structures and cannot be easily categorized. Hence to go beyond morphological features and take a complete view of all the states, quantitative methods are needed. The visual means are more computer intensive and for systems with more than two degrees of freedom, they may not be helpful. So we proposed the information entropy measure as a means to identify localized states [29]. Entropy is a well defined concept classically and an account of its applications in classical chaotic systems is given by Gutzwiller [1]. A useful review of general properties of entropy is given by Wehrl [31]. Powell and Percival [32] and Izrailev [33] have used information entropy measure for studying different properties of classical and quantum chaotic systems. Recently, it had also been used in the theory of quantum inference [34]. We employ the definition of information entropy of a quantum state as follows,

$$\mathcal{S}_n^{\alpha} = -\sum_{j=1}^{M} |\, a_{n,j}^{\alpha}|^2 \log |\, a_{n,j}^{\alpha}|^2, \tag{20}$$

where $a_{n,j}^{\alpha}$ are the eigenvector coefficients for the nth eigenfunction of the system which has M components, where M is the dimensionality of the Hamiltonian matrix for the problem. This measure is obviously basis-dependent. This measure identifies the localized states by showing a pronounced dip in the information entropy curve corresponding to various localized states. Figure 2 shows the information entropy plotted against logarithm of the state number for quartic oscillator for $\alpha = 90.00$, for the range of states from 950 to 2000. The well pronounced dips correspond to channel localized states. Shadow states are marked by less pronounced dips and are sometimes ambiguous. It must be emphasized that this measure picks out *only such localized states and no others*. We also computed the momentum space intensities, by taking the Fourier transform of the eigenfunctions. Their corresponding information entropies were also computed. We noticed that the same states showed minimum entropy in both momentum and configuration space, leading to the suspicion that these may be some sort of minimum uncertainty states. Further work in this direction is in progress.

The structure of the eigenvector coefficients, being a more fundamental entity, contains the signature of these localized states. The leading unperturbed basis vectors of the localized states have $(N,0),(N,2),(N,4)\ldots.$ structure, where N is an even integer depending on the particular eigenstate. It is seen that for localized states most eigenvector coefficients $a_{n,j}^{\alpha}$ are effectively zero and so very few basis states, in the above order, effectively contribute to the wavefunction.

It is now fairly well established from the pioneering work of Bohigas *et al* [35] that quantal energy level spacings distribution for classically chaotic quantum systems fall under three universality classes, determined by the symmetry properties of the Hamiltonian of the system. These level spacing distributions are modeled by appropriate ensembles from random matrix theory. For the Gaussian Orthogonal Ensemble (GOE), which models the fluctuation properties of the quartic oscillator, the eigenvectors are χ^2 distributed [36]. In this case, the RMT prediction for information entropy [37] is expressed in terms of digamma function Ψ as,

$$\mathcal{S}^{\mathrm{GOE}} \sim \log\left(\frac{D}{2}\right) - \frac{1}{D} - \Psi\left(\frac{3}{2}\right), \tag{21}$$

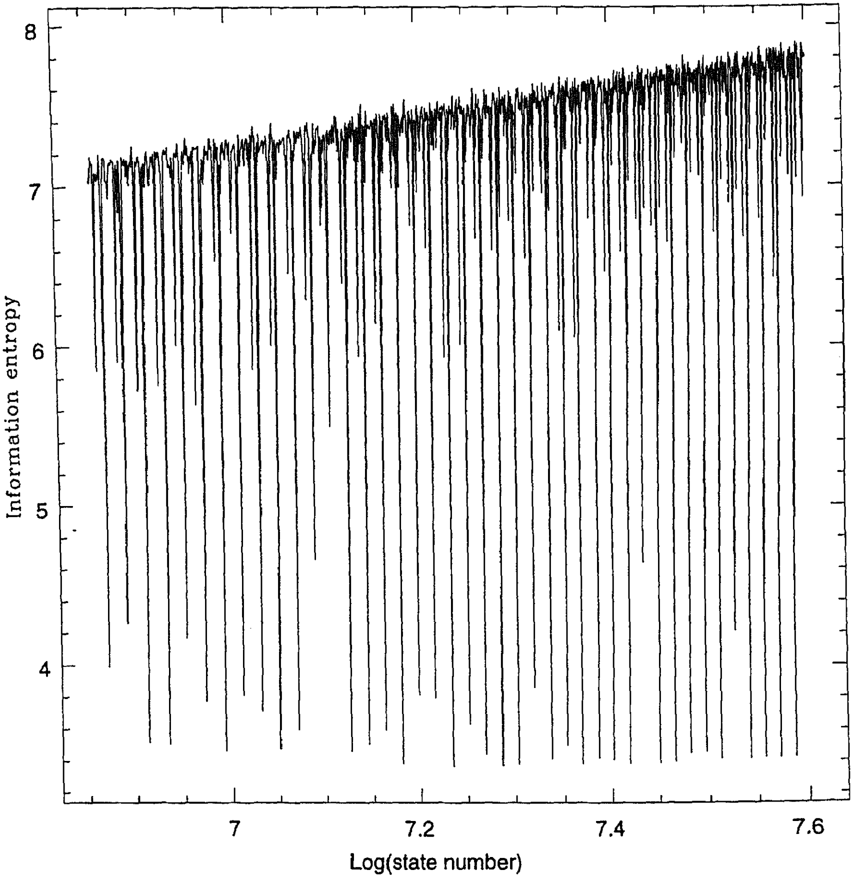


Figure 2. The information entropy plotted against the logarithm of the state number at $\alpha = 90.00$.

where D is large and is the dimensionality of the random matrices in the ensemble. In figure 2, the information entropy curve reveals an almost linear behaviour as predicted by the random matrix theory. In the unperturbed basis, the slope of the line fitted to the envelope of the information entropy curve is close to unity. However, we observed that in position and momentum basis the information entropy measure for the eigenfunctions does not exactly follow the RMT prediction. Thus, apart from basis dependence of eigenvector statistics, the question seems to hinge on the effective dimensionality of a state in the infinite dimensional Hilbert space. Integrated density of states could be one such measure of effective dimensionality. Another method is to employ the participation ratio defined as,

$$P_n = \frac{1}{\sum_{i=1}^{M} a_{ni}^4}$$

which gives the number of states that effectively participate in the building up of wavefunction [38]. We will be reporting more results on information entropy in § 7.

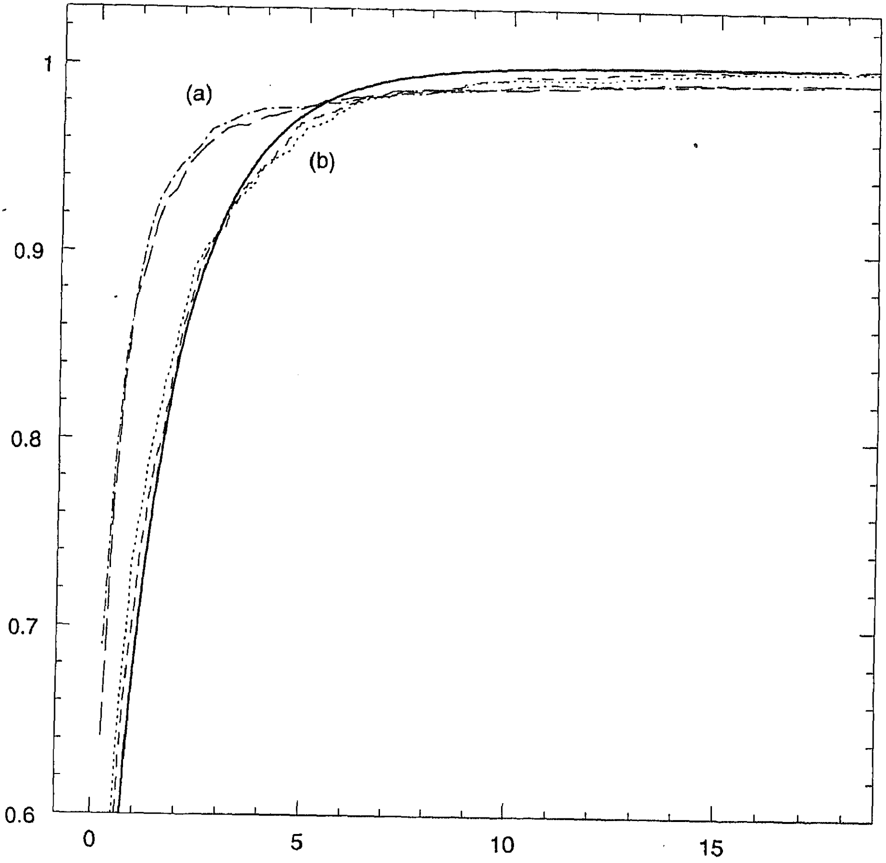


Figure 3. The cumulative eigenvector distribution curve. The solid line represents the standard cumulative Porter–Thomas distribution. The curves (**a**) correspond to two channel localized states, 409 and 423 at $\alpha = 64.66$. In (**b**), the dotted curve corresponds to a generic state and dashed curve corresponds to a strongly scarred state.

The eigenvector distribution for the Hamiltonian systems in infinite dimensional Hilbert space has not yet been well studied. There have been many studies for finite limensional systems and it has been noted that the Porter–Thomas distribution is ppropriate for time reversal symmetric chaotic systems [39]. Yet the basis dependence of ll such analysis is obvious. One way out is to select a generic and physical basis. In our ase the unperturbed basis is one such possibility; however it is clear that it preferentially reats the channel localized states.

The eigenvectors of the coupled quartic oscillator studied above follow the RMT rediction reasonably closely and three cases arise, namely, (i) the eigenvectors of those igenfunctions which are spread over all the classically accessible domain, (ii) the igenvectors of strongly scarred states and (iii) the eigenvectors of the channel localized tates. Figure 3 shows the cumulative Porter–Thomas distribution for the square of the eigenvector components normalized to unit mean given by

$$f(y) = 1 - \mathrm{Erf}\left(\sqrt{y/2}\right),$$

where Erf is the standard error function. It is evident that the cases (i) and (ii) have similar eigenvector distribution, which is close to the standard Porter–Thomas distribution (solid line). The strongly scarred state is state no. 552 at $\alpha = 64.66$ (see ref. [40] for a configuration space plot of state 552) , and the generic state is state no. 439 at the same value of α. In this sense, even the strongly scarred state can be said to exhibit random matrix like distribution. However the eigenvectors of the channel localized states deviate considerably from the RMT prediction, as figure 3 shows.

One reason for deviation from the RMT behaviour for the channel localized states could be that in the chosen basis states, the eigenvectors of the localized states have come in for a special treatment, which also makes these states stand out in the entropy analysis. A physical basis that may distinguish the scarred states from the rest could be some coherent basis set, but not the position or momentum basis as indeed we have already demonstrated that the minima of the entropy in these basis states also correspond only to channel localized states [29].

6. Adiabatic approaches for the quartic oscillator

In the last two decades, much progress has been made in the semiclassical quantization of classically chaotic systems and the work of Gutzwiller [1] in this direction has thrown more light on the classical-quantum correspondence. An adiabatic approach, using the standard semiclassical quantization, has been used to explain the channel and ring localized states and has been partially successful.

The basic premise of the adiabatic method is that the channel motion in two perpendicular directions have two vastly different time scales and hence slow and fast motions can be treated separately. For instance the motion along the channel is assumed to be slower than the motion perpendicular to it, resulting in the perpendicular action being an adiabatic invariant resulting in an adiabatic potential along the channel. Thus by separating motion in one direction from the other, the resulting locally integrable Hamiltonian is quantized to obtain the energy eigenvalues. However, this formalism does not separate the eigenvalues coming from different irreducible representations, quite like the quantum adiabatic method. For a fair comparison, the eigenvalues from this approach must be tallied with the average quantum eigenvalues from different irreducible representations. The adiabatic approximation works best for states localized far away from the origin for most of the time, like the channel localized states. This is due to the adiabatic curves having non-analytic properties at the origin. The quantum mechanically untamed potential $V(x, y; \alpha) = \alpha x^2 y^2$ has been studied in cartesian adiabatic approximation [41]. Ring type localization has been explained using the polar coordinate adiabatic breakup [42].

At this point some remarks on the ring states and the accuracy of adiabatic methods is in order. The adiabatic treatment involving semiclassical quantization, by its very nature, presupposes the existence of a series of particular type of scarred states. However, our own experience with the computation and visualization of nearly 2000 eigenfunctions for

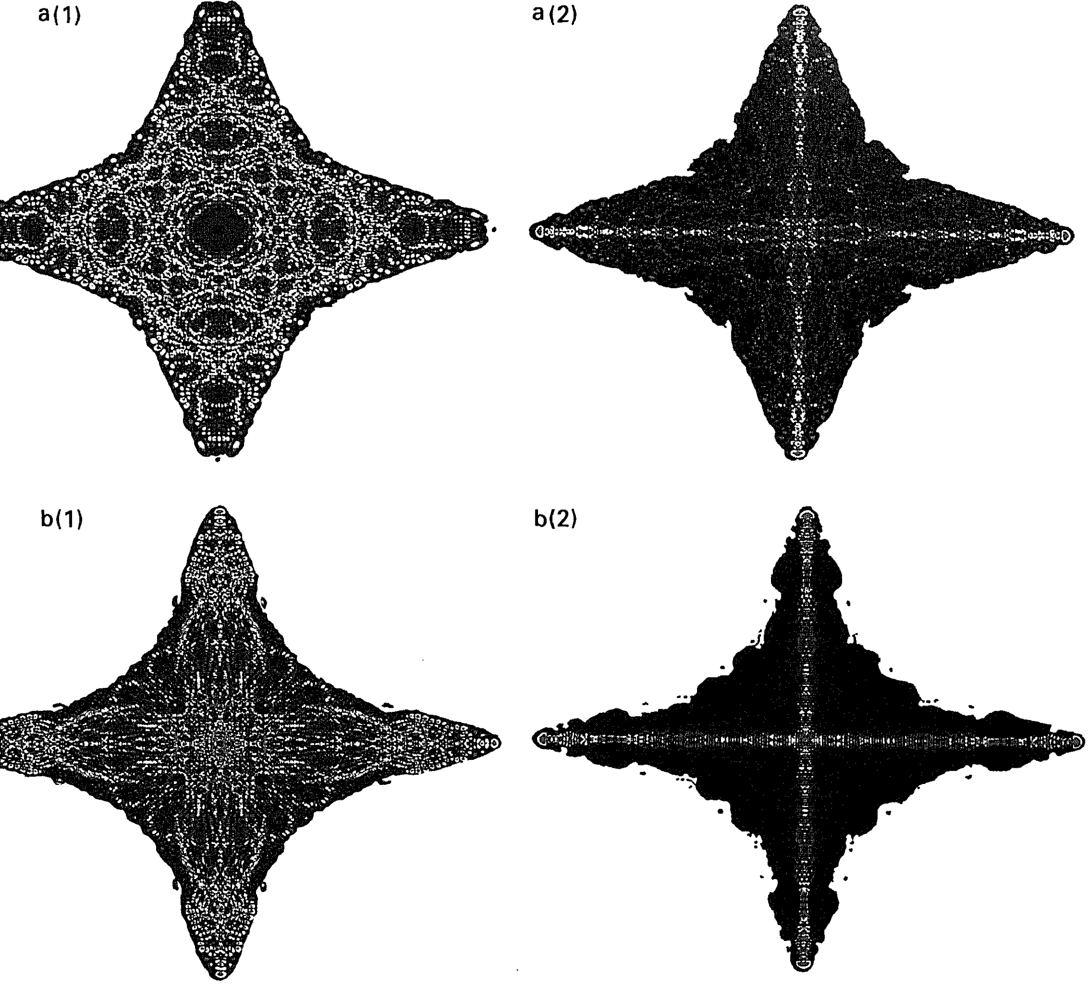


Figure 4. The configuration space intensities for two consecutive states for (a) $\alpha = 64.66$, states 1376 and 1377 with energies 2770.2953 and 2770.5087 respectively. (b) $\alpha = 90.00$, states 1857 and 1858 with energies 3642.9626 and 3643.0294 respectively. The colour code for all images is such that red is the maximum, blue is low and black is zero.

various α values in the deep semiclassical regime, amply confirms a series for channel states but does not show any evidence for a series of ring states. One reason for the absence of clean ring states in the deep semiclassical regime could be the instability of the underlying periodic orbits. It may be remarked that the previous studies [42] rely on low-lying eigenfunctions which do not show many avoided crossings and in spite of it being such a simple spectrum in the energy region considered, the identification of ring-states seems, at times, to be subjective. The highly excited region exhibits a large number of avoided crossings and many localized states stand on them and the adiabatic approach cannot unambiguously identify such states which differ in energy by less than the mean values of α, namely 64.66 and 90.00. Since the highly excited region of the spectrum has a large number of such cases, it becomes necessary to visualize eigenfunctions to confirm the results obtained by this approach. In this sense, the adiabatic approach provides more qualitative insight rather than quantitative result. As pointed out before, the visual methods have to complement these other techniques for correct interpretation of these approximate methods.

We proceed knowing fully well the limitations of this method. Eckhardt *et al* [28] have reported a variation of this method based on adiabatic stability. They have shown that one of the least unstable periodic orbits, namely the channel orbit, corresponds to a prominent peak in the Fourier transform of quantal spectra and have identified this orbit as the one that is responsible for scarring a series of eigenfunctions in quartic oscillator spectrum. Their analysis is based on the concept of adiabatic stability. Adiabatic stability refers to an orbit that is globally unstable as measured by its Lyapunov exponent but still stable for short times of the order its time period. The channel orbit, called the 'adiabatic well', is identified as the one around which the other orbits in the vicinity wind around. The adiabatic Hamiltonian associated with the channel orbit is constructed using these other orbits in a coordinate system defined by the channel orbit, which coincides with the cartesian coordinates in this case. Adiabatic eigenvalues are obtained by quantizing actions of these other orbits along and perpendicular to the channel orbit. The eigenvalues obtained using this formalism have a good correspondence with the quantal eigenvalues of the channel localized states. The adiabatically stable periodic orbits are thus shown to scar the channel states.

While the above works show that the existence of adiabatically stable channel orbits correlate with quantum density enhancements along the channels, this model ignores the effect of stability oscillations of the channel periodic orbits discussed earlier. We have noticed that the accuracy of the adiabatic approach depends on the stability of the channel orbits even in the case when the stable regions are of very small measure. Although the basic structure of these localized states remains the same, the degree of localization changes significantly.

The above semiclassical quantization schemes have largely ignored the discrete symmetries in the system. Sinha and Sheorey [43] have proposed a semiclassical quantization scheme that incorporates discrete symmetries using quantum input that takes cognisance of symmetries. In quartic oscillator the channel localized states are effectively one-dimensional states, and so a WKB-like scaling form for the irreducible representation Γ,

$$E(\Gamma,\alpha;N) = a_0 + a_1(N+\tfrac{1}{2})^{4/3} \tag{22}$$

has been applied for $(N,0)$ manifold, where N is any even integer for A_1 representation. With the data from $\alpha = 64.66$ case, it has been shown that the values a_0 and a_1 obtained from low-lying excited states predict the energies of the highly excited states in the region of state numbers from 400 to 500. The constant a_0 is a function of α. We have verified this relation for excited states extending up to 2000 states for different values of α. Again, there is a one-to-one correspondence between the predicted energy values and the quantum energy levels. In fact the adiabatic approach also predicts a formula very similar to the above one and is capable of giving a functional form to a_0. More detailed results on this approach will be published elsewhere.

It should again be emphasized that only a subclass of scarred states, the channel localized states and ring states have thus far been amenable to any kind of approximate description. The plethora of various structures exhibited by eigenfunctions of quartic oscillators still remain a mystery. As of today, another outstanding problem is the construction of adiabatic wavefunctions that would resemble the actual quantum wavefunctions.

7. Localization and stability of channel orbits

Since it is well known that the underlying periodic orbits play a crucial role in the semiclassical localization, the next aspect to be studied is the effect of bifurcation and the stability of these orbits on localization. The bounded stability oscillations allow us to determine the values of α at which the channel orbit undergoes bifurcations. The value $\alpha = 90.00$ is one such point at which the channel orbit is about to lose stability through a pitchfork bifurcation. We computed the eigenvalues and eigenfunctions for the first 2000 states of the quartic oscillator for values of α ranging from 88.00 to 96.00.

The information entropy in the unperturbed basis states for all these eigenfunctions was computed. The preliminary results are indicative of a correlation between the stability of the channel orbit and degree of eigenfunction localization. Higher the stability of the channel orbit, the localization is more compact and clean. For instance, figure 5 shows a series of highly excited eigenfunctions in configuration and unperturbed space for various values of α. It is clear from this figure that comparing the eigenfunctions at $\alpha = 64.66$ and $\alpha = 90.00$, the latter is highly compact and almost collapses on the underlying periodic orbit. Again the structure of unperturbed eigenfunction for $\alpha = 90.00$, namely the eigenvector coefficients reveal that most of the components are almost zero, except for a few dominant ones which go up to build the wavefunction. However at $\alpha = 64.66$, the eigenfunction in unperturbed space is spread over many more basis states than for $\alpha = 90.00$.

The information entropy measure also reflects this trend as we go across $\alpha = 90.00$. We calculated averaged information entropy for a particular α value by taking the mean of information entropies,

$$\langle S_\alpha \rangle = \frac{\sum_\sigma S_\sigma^\alpha}{m}$$

for a group of m localized states represented by σ within some energy range. Strictly speaking this quantity may not be physically meaningful, but it is a coarse-grained value that would reveal the average trend in the degree of localization. The participation ratio also shows an identical trend. Figure 6 shows the average entropy plotted against α and it has a clean dip at $\alpha = 91.00$. This is when the channel orbit has just begun to lose stability. The dip at $\alpha = 91.00$ seems to justify the remark that the influence of a new orbit on an eigenstate is not clear immediately after the bifurcation [44]. A careful analysis is still needed to uncover the changes in eigenfunction structures that accompany bifurcations of the underlying periodic orbits. This could also mean that though the

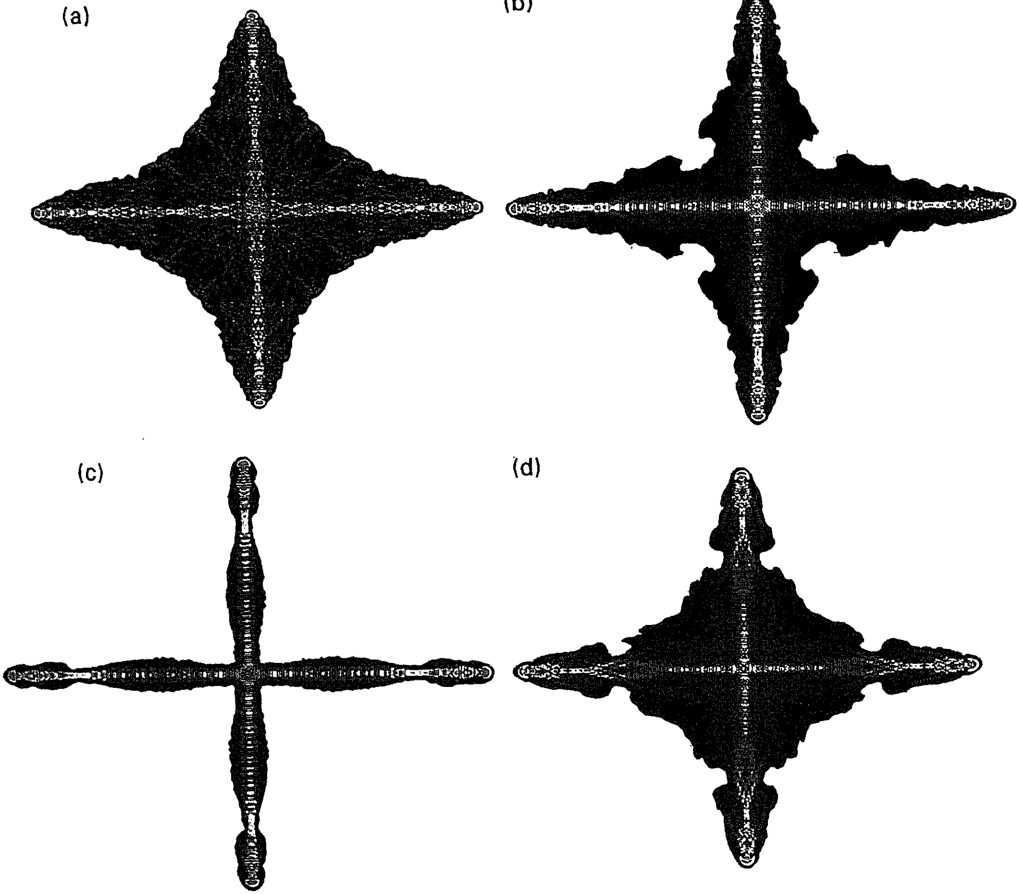


Figure 5. Configuration space intensities for four localized states for (a) α=64.66, (b) 88.00, (c) 90.00 and (d) 96.00, all with the same principal unperturbed state (252, 0). The colour code for all the images is such that red is the maximum, blue is low and black is zero.

channel orbit plays an important role in channel localization, the contribution of the other orbits in the vicinity cannot be ignored.

All these results in totality imply a strong correlation between the stability of channel orbits and degree of localization, and also are indicative of the dominant role played by the channel orbit although other nearby orbits do contribute to localization along the channel. In a similar spirit, though not quite the same, Atkins and Ezra [44] have qualitatively studied the effect of pitchfork bifurcations of the diagonal classical periodic orbit on the eigenstates of the quartic oscillator. This orbit is always unstable beyond $\alpha = 6.0$. The bifurcation of the orbit is shown to be associated with the spread in the eigenstate in configuration space transverse to the 45° orbit considered. However, they have confined themselves to values of α up to 15 in the low-energy spectrum only.

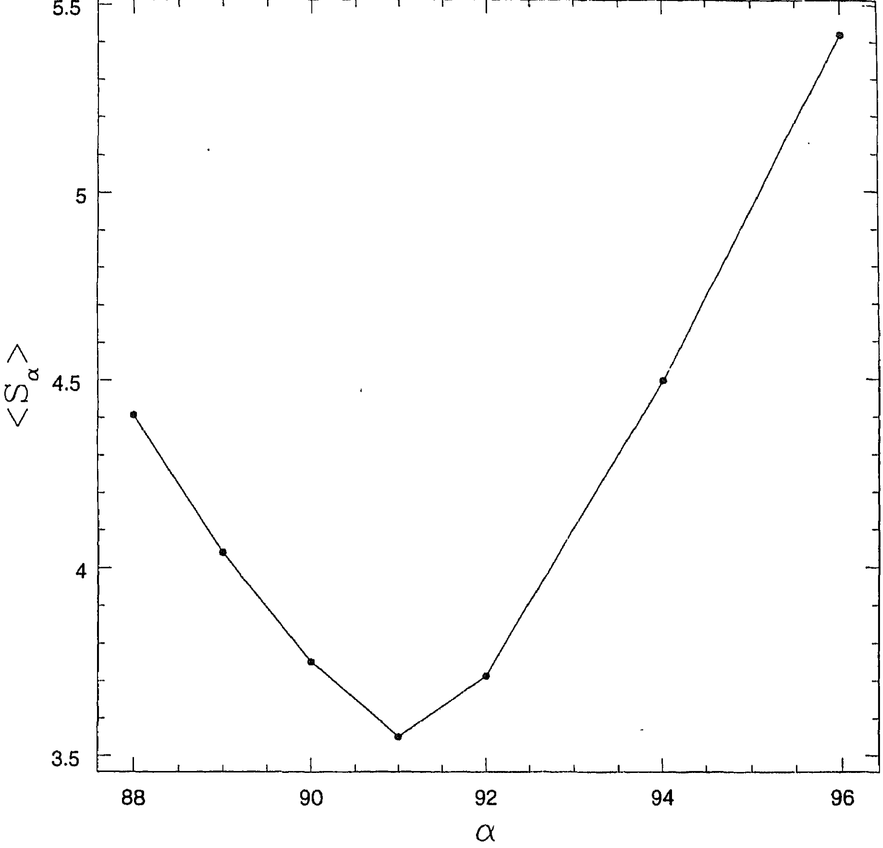


Figure 6. The plot of average information entropy as a function of α.

8. Conclusions

A comprehensive theory encompassing all the nuances of localization in quantum mechanics is still an open problem. Our extensive studies on the eigenfunctions of the two dimensional quartic oscillator potential shows the existence of various scarred states, in particular the channel localized states at least up to about 2000 levels from the ground state. We expect that these will continue to exist throughout the spectrum. The channel localized states alone can be identified by the information entropy method, or other measures of localization such as the participation ratio.

We have studied localization in a time independent classically chaotic system, and is quite distinct from the localization in kicked systems such as the rotor. The regions of parameters we have studied the oscillator in are ones of extreme chaos, although an important periodic orbit scarring the set of localized states we study oscillates in stability. We have numerically studied the effect of these oscillations and concomitant classical bifurcations on eigenfunction localization, and found that such measures as information

Experiments on quantum chaos using microwave cavities: Results for the pseudo-integrable L-billiard

A KUDROLLI and S SRIDHAR
Physics Department, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

Abstract. We describe microwave experiments used to study billiard geometries as model problems of non-integrability in quantum or wave mechanics. The experiments can study arbitrary 2-D geometries, including chaotic and even disordered billiards. Detailed results on an L-shaped pseudo-integrable billiard are discussed as an example. The eigenvalue statistics are well-described by empirical formulae incorporating the fraction of phase space that is non-integrable. The eigenfunctions are directly measured, and their statistical properties are shown to be influenced by non-isolated periodic orbits, similar to that for the chaotic Sinai billiard. These periodic orbits are directly observed in the Fourier transform of the eigenvalue spectrum.

Keywords. Quantum chaos; billiards; eigenvalues; eigenfunctions; random matrix theory; pseudo-integrable.

PACS Nos 05.45; 03.65; 71.55

1. Introduction

The issue of non-integrability in quantum mechanics is relevant in a variety of physical systems. Indeed initial developments were in nuclear physics, where random matrix theory (RMT) was developed to account for the statistical properties of the eigenvalue spectra of heavy nuclei [1, 2]. Similar and hence universal behavior has been shown numerically in low dimensional billiard systems [3]. This universal behavior is not just confined to quantum mechanical systems such as atoms, electrons and nuclei, but is also seen in wave mechanical systems like acoustic and electromagnetic systems. The common feature in all these apparently disparate systems is that the corresponding ray or particle dynamics is chaotic. Significant developments in semiclassical theory have shown the importance of classical structures, such as periodic orbits, in organizing quantum behavior [4–7]. Although many experimental systems exhibit quantum or wave chaos, few of these yield properties which are directly comparable to analytical results. Hence theoretical developments have relied on numerical simulations to support their claims. Although computational tools are rapidly increasing in their applicability, there are nevertheless many important problems which are still not amenable to reliable numerical solutions. In this paper we describe an experimental approach which leads to important insights to the role of classical chaos and even disorder in quantum or wave mechanics. This approach, which utilizes microwave cavities, is best suited to the study of "billiard" problems that examine the quantum properties of a particle in model 2-D geometries. In the past five years these experiments have resulted in several important contributions to the field of quantum chaos. The experiments have led to the first direct observation of scarred eigenfunctions [8]. The statistical properties of the eigenfunctions for the chaotic Sinai-stadium geometry were seen to be in good agreement with the universal Gaussian orthogonal ensemble (GOE) of RMT [9]. It was further found that chaotic geometries with non-isolated periodic orbits like the Sinai billiard show deviation from universal behavior. The eigenfunctions of disordered geometries (cavities with many finite sized scatterers) have shown the effects of localization on their statistical properties [9]. This has stimulated interesting theoretical work on the spatial distribution of eigenfunctions using supersymmetric non-linear σ-models [10–12] and to date remain the only experiments in which these theories can be tested. The eigenvalue statistics have also been analyzed and shown to be in direct agreement with GOE for the Sinai-stadium [13]. The experiments have also led to the experimental verification [14] of the theorem of isospectral domains [15] and resolved an important issue of degeneracies. Periodic orbits were observed in both closed and open billiard systems by Fourier transforming the eigenvalue and transmission spectrum respectively [16, 17]. A consistent theme of this work has been the gleaning of universal behavior and deviations from such universal behavior, by exploiting the ability to study controlled variations in geometry afforded by the microwave experiments. In this paper, we discuss in detail the eigenvalue and eigenfunction properties of an important class of billiards which are pseudo-integrable. This example is important because it is an intermediate system that is neither integrable nor fully chaotic. The theoretical description of intermediate systems is challenging and it is hoped that these quantitative experimental results will lead to rigorous tests for theory. We begin with an introduction of the experimental techniques.

2. Experimental techniques

The experiments utilize "thin" microwave cavities, in which below a cutoff frequency, the 2-D scalar limit of the Maxwell–Helmholtz wave equation applies, and the correspondence to the Schrödinger equation is exact. For a cavity of arbitrary cross-section in the $x-y$ plane but uniform along the z- direction, the z-component of the wave vector is quantized as $k_z = p\pi/d$, where p is an integer and d is the cavity thickness along the z- axis. Maxwell's equations then reduce to

$$[\nabla^2 + (k^2 - (p\pi/d)^2]\{E_z, B_z\} = 0, \tag{1}$$

where ∇ is the 2-D Laplacian operator. Of course the EM field is a vector field, however a special case occurs for the transverse magnetic (TM) modes for which $B_z = 0$, and further when $p = 0$. In this limit, (1) reduces to the Schrödinger equation $(\nabla^2 + k^2)$ $\Psi = 0$, where $\Psi \longleftrightarrow E_z$. This limit can be experimentally achieved by confining measurements to the frequency range $f < f_c = c/d$. For $d = 6\,\mathrm{mm}$, the cutoff frequency $f_c = 25\,\mathrm{GHz}$, and for all lower frequencies, the correspondence between the Schrödinger and Maxwell wave equations is exact in the thin 2-D cavities. Cavities are fabricated with the 2-dimensional cross-section cut out from 6 mm thick copper sheets, and placed between two copper plates. Coupling to the cavities is accomplished by loops terminating coaxial cables – the loops couple to the microwave magnetic fields at the perimeter. The cavity response is examined using an HP 8510 Network Analyzer (ANA). The eigenvalue spectrum is obtained from the resonances observed in the transmission spectrum, with the energy eigenvalues E_n obtained from the resonant eigenfrequencies f_n using $E_n = f_n^2$. By using appropriate care, in particular coupling at several locations, it is possible to observe the first 1000 or so levels of the spectrum, without any missed levels. Thus although the finite conductivity of the copper walls leads to broadened resonances, it is not necessary to use superconducting cavities in order to carry out reliable experiments on eigenvalue statistics. One of the important and unique features of the experiments is the ability to directly measure eigenfunctions. This method, first devised by one of us [8, 18], utilizes the perturbation of the cavity resonance by a small metallic ball. Placing the ball at coordinates (x,y) in the cavity leads to a shift of the resonance frequency $\Delta f_n(x,y) = -\beta\Psi_n^2(x,y)$, from which the eigenfunction Ψ_n can be directly determined by measuring the frequency shift $\Delta f_n(x,y)$ using the ANA. Further details about the method can be found in [18].

3. Pseudo-integrable geometries

An important class of billiard systems are the pseudo-integrable geometries as they are the closest step away from integrable systems. Most polygonal billiards, with the exception of rectangles, 30°–60°–90° angle triangles, and the equilateral triangle are pseudo-integrable. The non-integrability arises at the vertex, when it is not 90°, 60°, 45°, 30°. Trajectories diverge at this point and lead the classical phase to be a surface of genus greater than one. Since the classical phase space has to be a torus for the motion to be integrable, a higher genus surface ensures non-integrability. A formula for relating the genus, g, to the angles of the vertex has been derived and is given as [19, 20]:

$$g = 1 + \frac{N}{2}\sum\left[\frac{m_i + 1}{n_i}\right], \tag{2}$$

where n_i and m_i are such that $\alpha_i = m_i\pi/n_i$ are the internal vertex angles, and N is the smallest of the integer set n_i. A surface of genus 2 which corresponds to a pseudo-integrable system is the L-shaped billliard [21, 22]. However there are similarities to integrable systems, as the periodic orbits in these geometries also occur in families, which bifurcate at the vertex which is not-integrable. They also appear to have the same asymptotic proliferation of periodic orbits [20]. However numerical work has shown that the eigenvalue statistics of these systems show non-Poisson statistics. Below we shall discuss results on a well-known example, the L-shaped billiard.

3.1 *Eigenvalue statistics*

The eigenvalue statistics of a pseudo-integrable (PI) L-shaped cavity was studied with a 44 cm by 21.8 cm copper cavity with a 9.55 cm by 9.65 cm copper piece at one of the corners. In the frequency range 0.045 GHz to 18.497 GHz, 1000 energy levels were obtained. This data also shows good agreement with the Weyl formula [23]. The nearest neighbor energy spacing $P(s)$, and spectral rigidity $\Delta_3(L)$ were studied to understand the

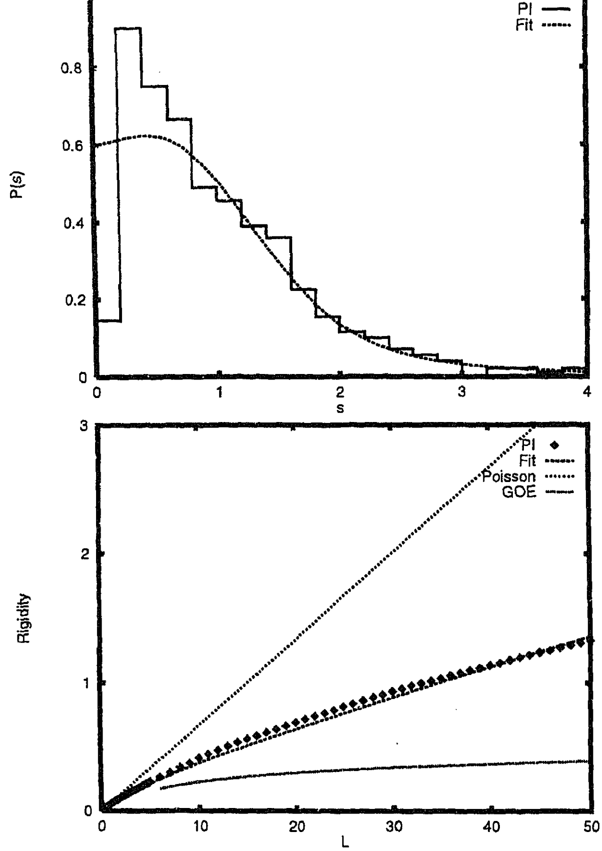


Figure 1. Eigenvalue statistics for the pseudo-integrable (PI) geometry. (Top) $P(s)$ for the pseudo-integrable geometry. The interpolation to fit the data is done using eq. (3) with $\beta = 0.62$. (Bottom) The rigidity, $\Delta_3(L)$ versus L for the pseudo-integrable geometry. The interpolation is done using (4) with $\beta = 0.65$.

nature of the eigenvalue spectrum. Even though the classical phase space structure is closer to that of an integrable geometry, level repulsion is quite evident in figure 1. The spectral rigidity $\Delta_3(L)$ was also analyzed and is also shown in figure 1. The resulting data lie intermediate between the universal integrable and chaotic curves for both the statistical measures in figure 1. We have shown elsewhere [13] that the auto-correlation function [24] has a correlation hole, as it goes below the asymptotic value in the time scales comparable to the inverse of the mean energy spacing. However, the effect is not as strong as in the chaotic case, and hence in this respect also, pseudo-integrable geometries display intermediate behavior.

For intermediate cases, Berry and Robnik [25] proposed a theory that the intermediate case can be characterized by the volume of phase space which is chaotic. While this argument was for quasi-integrable cases that have a mixed phase space, we have attempted to use it to describe the pseudo-integrable data.

$$P(s) = \beta \exp(-s) + (1-\beta)\frac{\pi}{2} s \exp\left(-\frac{\pi}{4}s^2\right), \tag{3}$$

where, β is the parameter which characterize the ratio of phase space which is integrable. The fit corresponding to eq. (3) is displayed in figure 1, where the data is well described by the fit for s greater than the first few bins. The β value obtained is 0.62. While there are a few missing levels, the discrepancy from (3) may be due to poor statistics, but is more likely to be real, indicating that a more refined theory is required. The experimental spectral rigidity in figure 1 is compared with theory [25, 26]:

$$\Delta_3(L) = (L/15)\beta + (1-\beta) \times \left(\frac{1}{\pi^2}\ln(L) - 0.007\right). \tag{4}$$

The β value which approximates the experimental curve is 0.65, in reasonable agreement with that obtained using (1). The auto correlation function for the intermediate case was also derived by Alhassid and Whelan [24] using the same argument of Berry and Robnik. The data for the L-billiard again fits the expression derived by them, expression for $\beta = 0.65$ [13]. Therefore a β of approximately 0.65 roughly describes the intermediate nature of the data for the three different statistical measures. But such matching should be taken with caution since the phase space structure of pseudo-integrable systems is different from that of a quasi-integrable system.

3.2 *Eigenfunction properties*

As noted earlier, one of the unique features of the experiments is the ability to obtain directly the eigenfunctions. While there are at least a few studies of the eigenvalues of these systems [19, 27, 28], the eigenfunctions have not been studied in much detail. Below we discuss several unique features of the eigenfunctions of the L-billiard. Eigenfunction data were obtained in the frequency range 0.5 GHz and 4.5 GHz, and can be easily taken into a higher frequency range. Sample eigenfunctions are displayed in figure 2, and appear to have similar structure as the eigenfunctions of the Sinai billiard in the same frequency range. The averaged density distribution is shown in figure 3 for

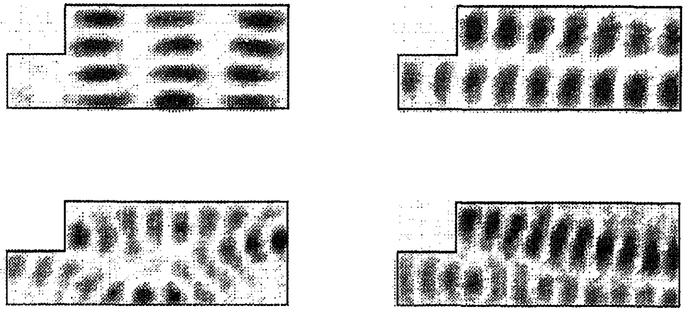


Figure 2. Sample experimental eigenfunctions of the pseudo-integrable system. The corresponding frequencies are 3.032 GHz, 3.346 GHz, 4.079 GHz, and 4.462 GHz. Dark regions correspond to high amplitudes.

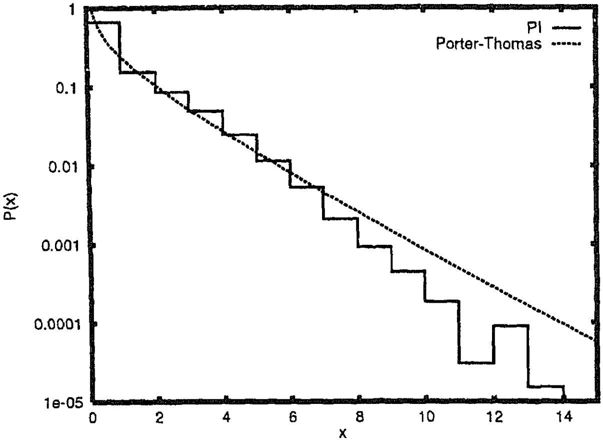


Figure 3. Density distribution of the eigenfunctions of the pseudo-integrable system ($x = |\psi|^2$).

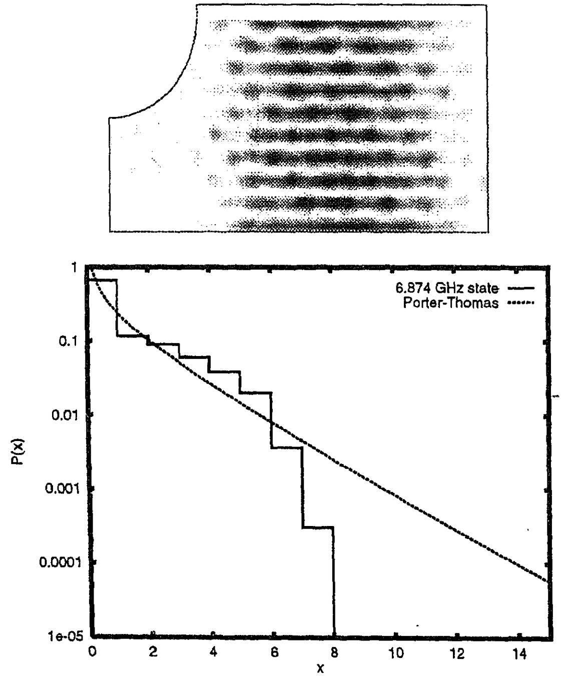


Figure 4. (Top) A bouncing ball state of the quarter Sinai billiard, and (Bottom) its density distribution.

igenfunctions between 3.0 GHz and 4.5 GHz. A useful statistical measure to characterize igenfunctions is the density distribution, $P(|\psi|^2)$, which gives the probability of finding a articular $|\psi|^2$ in the cavity. The distribution corresponding to the L-billiard is istinctively different from an integrable rectangle, and in fact is closer to that of

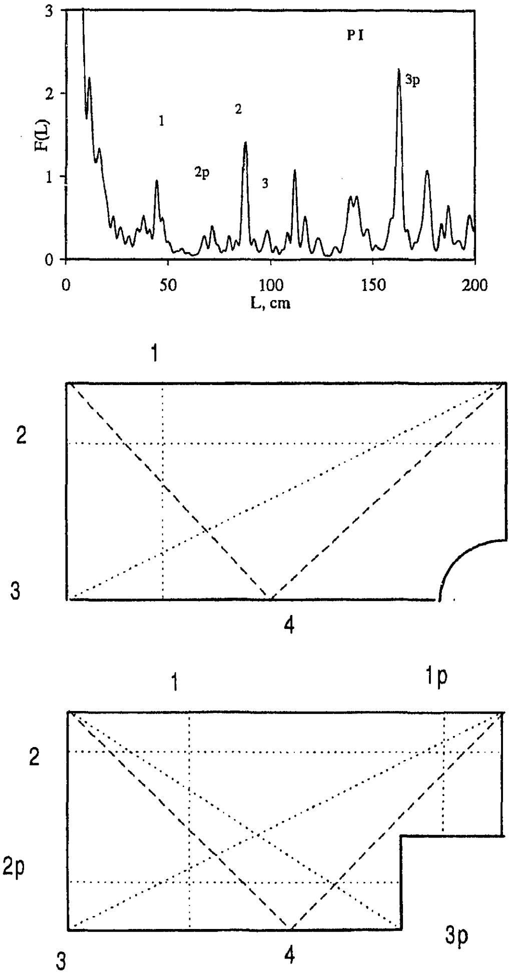


Figure 5. (Top) Fourier transform of the eigenvalue spectrum of the L-billiard. (Bottom) The peaks which can be associated with periodic orbits are labelled and shown. Also shown (middle) are corresponding orbits for the Sinai billiard.

chaotic geometries. For a rectangle, $P(|\Psi|^2)$ is truncated at $|\Psi|^2 = 4$. Thus large intensities are forbidden in the rectangle, in contrast to chaotic systems which obey the Porter–Thomas distribution [2] (see figure 3) that shows a finite though exponentially vanishing probability of finding large amplitudes. We have demonstrated this for the chaotic Sinai-stadium geometry. Closer examination shows that the data for the L-billiard actually disagrees with P-T at high densities, and instead are in closer agreement with the Sinai billiard than with the Sinai-stadium. The similarity of the behavior of the pseudo-integrable system with the Sinai billiard should not be surprising, as both have a similar structure of non-isolated periodic orbits. Very little work on the eigenfunction properties has been done for pseudo-integrable systems [29], and new theoretical developments will have to take place to describe experimental data.

3.3 *Deviations due to non-isolated periodic orbits in the Sinai billiard and the L-billiard*

As discussed in the previous section, some states of geometries like the Sinai and stadium billiards and the L-billiard, have structures which are far from irregular. In fact some show eigenfunctions that have rectangular structure, similar to the eigenfunctions of a rectangle. These eigenfunctions have an amplitude distribution with far less probability at large amplitudes. Figure 4 shows an example of a bouncing ball state of the Sinai billiard with $|\psi|^2$ almost evenly distributed between the parallel sides. The corresponding $P(|\psi|^2)$ is also displayed in the same figure that shows strongly truncated distribution. So when eigenfunctions are averaged over all states, some that are the bouncing ball, one gets an over all departure from Porter–Thomas. Observation of these small deviations clearly requires very sensitive experiments such as presented here. Recent experiments [30] with stadium cavities which probed the distribution of widths could not see this deviation from universality due to lack of accuracy and statistics [31].

3.4 *Fourier transforms and direct observation of periodic orbits*

A very useful means to directly observe and even measure the shortest periodic orbits (PO) is by taking the Fourier transform (FT) of the measured eigenvalue spectrum. The FT is shown in figure 5, and has several peaks, which correspond very well with the PO shown in figure 5. Note that only the non-isolated PO, i.e. bouncing ball orbits, are clearly present. The strength of a given PO in the FT is likely to be related to its area of stability [32, 22]. It appears that in order to manifest themselves in the FT of the eigenvalue spectrum, the relevant PO must have sufficient phase space areas of stability. Experiments on the Sinai billiard also show similar peaks in the FT since the non-isolated PO are similar in both cases (see figure 5).

4. Conclusion

The experiments described in this paper represent a unique and novel approach to studying model problems in quantum or wave chaos. We have examined the behavior of a system that is intermediate between integrable and chaotic. The statistical properties of the eigenvalues and eigenfunctions show intermediate behavior between Poisson and the Gaussian orthonormal ensemble of random matrix theory. Comparisons with existing models which examine the quasi-integrable systems shows reasonable agreement but also point to a need for a more refined theory for pseudo-integrable systems. Our results demonstrate that powerful insights can be obtained by a judicious combination of experiment and comparison with theory.

References

[1] M L Mehta, *Random matrices*, 2nd ed. (Academic, New York, 1990)
[2] C Porter, *Statistical theories of spectra: Fluctuations* (Academic Press, 1965)
[3] O Bohigas, M J Giannoni and C Schmit, *Phys. Rev. Lett.* **52**, 1 (1984)
[4] M Gutzwiller, *J. Math. Phys.* **12**, 343 (1971)

[5] M Gutzwiller, *Chaos in classical and quantum mechanics* (Springer-Verlag, New York, 1990)
[6] M V Berry and M Tabor, *Proc. R. Soc. London* **A356**, 375 (1977)
[7] E J Heller, *Phys. Rev. Lett.* **53**, 1515 (1984)
[8] S Sridhar, *Phys. Rev. Lett.* **67**, 785 (1991)
[9] A Kudrolli, V Kidambi and S Sridhar, *Phys. Rev. Lett.* **75**, 822 (1995)
[10] Y V Fyodorov and A D Mirlin, *Phys. Rev.* **B51**, 13403 (1995)
[11] V Falko and K Efetov, unpublished
[12] V Prigodin, N Taniguchi, A Kudrolli, V Kidambi and S Sridhar, *Phys. Rev. Lett.* **75**, 2392 (1995)
[13] A Kudrolli, S Sridhar, A Pandey and R Ramaswamy, *Phys. Rev.* **E49**, R11 (1994)
[14] S Sridhar and A Kudrolli, *Phys. Rev. Lett.* **72**, 2175 (1994)
[15] C Gordon, D Webb and S Wolpert, *Bull. Am. Math. Soc.* **27**, 134 (1992)
[16] A Kudrolli, Ph.D. thesis, *Experiments on quantum and electromagnetic chaos*, Northeastern University (1995)
[17] A Kudrolli and S Sridhar, in *Proceedings of the 4th Drexel Conference on Quantum-Nonintegrability* (World Scientific, Singapore, 1996)
[18] S Sridhar, D Hogenboom and A Kudrolli, in *Quantum non-integrability* edited by D H Feng and J M Yuan (Gordon & Breach, 1993)
[19] P J Richens and M V Berry, *Physica* **D2**, 495 (1981)
[20] E Gutkin, *Physica* **D19**, 311 (1986)
[21] J Vega, T Uzer and J Ford, *Phys. Rev.* **E48**, 3414 (1993)
J Vega, T Uzer and J Ford, *Phys. Rev.* **E52**, 1490 (1995)
[22] D Biswas, (preprint)
[23] H P Baltes and E R Hilf, *Spectra of finite systems* (B-I Wissenschaftsverlag, Mannheim, 1978)
[24] Y Alhassid and N Whelan, *Phys. Rev. Lett.* **70**, 572 (1993)
[25] M V Berry and M Robnick, *J. Phys.* **A17**, 2413 (1984)
[26] M V Berry, *Proc. R. Soc. London* **A400**, 229 (1985)
[27] F Haake, G Lenz, P Seba, J Stein, H-J Stockman and K Zyczkowski, *Phys. Rev.* **A44**, R6161 (1991)
A Shubo and Y Shimizu, *Phys. Rev.* **E47**, 54 (1993)
[28] R L Weaver, *J. Acoust. Soc. Am.* **85**, 1005 (1989)
[29] P Bellomo and T Uzer, *Phys. Rev.* **E50**, 1886 (1994)
[30] H Alt, H-D Graf, H L Harney, R Hofferbert, H Lengeler, A Richter, P Schardt and H A Weidenmuller, *Phys. Rev. Lett.* **74**, 62 (1995)
[31] A Kudrolli and S Sridhar, *Phys. Rev. Lett.* **76**, 3036 (1996)
[32] D Biswas and S Sinha, *Phys. Rev. Lett.* **70**, 916 (1993)

Quantum chaology: The photoeffect and beyond

BALA SUNDARAM
Department of Physics, University of Texas at Austin, Austin, Texas 78712, USA
Present address: Department of Mathematics, CSI-CUNY, 2800, Victory Boulevard, Staten Island, NY 10314, USA

Abstract. I will show how aspects of quantum chaology are relevant even in a seemingly well understood quantum phenomenon like the photoelectric effect. This example together with recent experiments in atom optics are used to define and discuss the larger questions, recent progress made by us in resolving some of these issues and future directions.

Keywords. Quantum chaos; photoelectric effect; atom optics; standard map.

PACS Nos 03.65; 05.45; 32.80

1. Introduction

In his 1987 Bakerian lecture to the Royal Society in London, M V Berry defined quantum chaology as "the study of semiclassical, but nonclassical, behaviour characteristic of systems whose classical motion exhibits chaos" [1]. For more than two decades prior to this, work on systems which exhibited such behaviour had been ongoing under the nominal definition of 'quantum chaos' [2]. Unfortunately, the very nature of quantum mechanics precludes many of the defining characteristics of classical chaos, making this latter definition burdensome. In fact it has even led some researchers to ignore the unquestionable fact that there are quantifiable changes in the quantum dynamics when the limiting classical dynamics exhibits chaos, as compared with a nonchaotic counterpart.

In this article, I shall illustrate that aspects of these changes are readily observed even in a simple quantum paradigm – the photoelectric effect. The fact that this phenomenon embodies wave-particle duality makes it ideal to introduce and discuss quantum-classical correspondence in the larger context of non-integrable systems. I will then consider another new experimental system [3] in atom optics which facilitates the realization of theoretical paradigms like the quantum δ-kicked rotor and provides an example of a different manifestation of quantum chaology. Proceeding from these illustrations, I shall define the larger picture in the study of quantum dynamics with classically chaotic limit and briefly discuss our work, in collaboration with Rainer Scharf, in integrating all the relevant issues within the framework of a single model system. Concluding remarks address ongoing directions with a view on future experiments in atom optics.

Most of us remember the photoelectric effect as the emission of electrons when a surface is irradiated with light. The threshold is defined by the condition $h\nu = W$ where ν is the frequency of the light and W a characteristic binding energy for the electron. It was soon realized that a generalized multiphoton threshold $N_{\rm th}h\nu = W$ was also possible though the intensity of the light had to be considerably higher with increasing $N_{\rm th}$. To put this principle to practice required the advent of the laser. Atoms and molecules became the targets of choice, W was taken to be the ionization or dissociation energy and the age of experimental multiphoton physics was born. Since that time, laser power has rapidly gone through the kilo-, mega-, giga- and tera- prefixes and now borders on the peta-Watt regime. This increase in intensity has brought with it new and manifestly non-perturbative phenomena such as above-threshold ionization (where the electron absorbs more photons than it needs just for the ionization threshold) and harmonic generation (where the irradiated system behaves as a driven anharmonic oscillator and scatters photons at frequencies which are large multiples of the incident frequency). These have led to a revision of the traditional analysis of multiphoton physics [4]. Most importantly from our perspective, it has brought dynamics to the foreground.

Despite these changes in our understanding of the photoeffect, the one feature which had not been challenged was the fact that the probability of freeing the electron was still expected to *increase* with *increasing* intensity of the light. In other words, *the lifetime of the quantum state was expected to decrease* with growing intensity.

The phenomenon of 'stabilization' was reported in numerical experiments [5] which considered a model Hamiltonian, for a 1-D atom, of the form

$$H(x,p,t) = p^2/2 - 1/\sqrt{1+x^2} + xF\cos(\omega t + \phi) \text{ , (a.u.)} \qquad (1)$$

where F, ω, and ϕ are the field strength, frequency, and phase of the oscillating electric field. This 1-D model potential asymptotes to the Coulombic potential for large x, but eliminates the singular behaviour at the origin. On solving the corresponding time-dependent Schrödinger equation, it was found that on *increasing F, the probability for freeing the electron decreased dramatically*, for fixed interaction time. The wavefunction at this time also exhibited peaks which were consistent with an effective double well potential [5]. Both 'stabilization' and the distinct 'dichotomous' form of the residual wavefunction were soon confirmed in more realistic 3-D simulations [6]. We shall show how stabilization provides an illustration of effects in 'quantum chaology' which can then be used to predict the extent of stabilization and the shape of the wavefunction [7].

It is true that advances in classical nonlinear dynamics and chaos have been successfully applied to strongly coupled quantum systems like the photo-absorption spectrum of Rydberg atoms in strong magnetic fields [8], microwave ionization of highly excited hydrogen atoms [9], and the excitation of doubly excited states of helium atoms [10]. However, the fact that the atoms are initially in the ground state, far from the classical-quantum correspondence limit, makes the application of these ideas to the remarkable phenomenon of 'stabilization' in super-intense laser fields still more counter-intuitive [11].

2. Stabilization and scarring of wavefunctions

In the high-field limit ($F > 1$ atomic unit meaning that it is greater than the binding potential) the smoothed Coulomb potential in (1) can be treated as a perturbation on the regular, classical motion of a free electron in an oscillating field. So, let us first consider the Hamiltonian for the one-dimensional motion of a free electron in the oscillating electric field $E = F\cos(\omega t \phi)$,

$$H_0(x, v, t) = v^2/2 - xF\cos(\omega t + \phi) \ . \tag{2}$$

The classical equations of motion can be integrated exactly and the solution for the position of the electron as a function of time

$$x(t) = x_0 - \frac{F}{\omega^2}[\cos(\omega t + \phi) - \cos\phi] + \left[v_0 - \frac{F}{\omega}\sin\phi\right]t, \tag{3}$$

describes a particle that oscillates back and forth in the electric field with frequency $\nu = \omega/2\pi$ and amplitude $\alpha = F/\omega^2$, and drifts away from its initial position, x_0, with velocity

$$v_\mathrm{d} = v_0 - \frac{F}{\omega}\sin\phi \ . \tag{4}$$

(Note that even if the initial velocity, v_0, is zero, a nonzero initial phase, ϕ of the oscillating field can lead to a large drift velocity, $v_\mathrm{d} = -(F/\omega)\sin\phi$.) The classical motion is considerably simplified if we consider variables in the oscillating frame

$$\begin{aligned} q &= x + \alpha[\cos(\omega t + \phi) - \cos\phi] \\ p &= v - \left[\frac{F}{\omega}\sin(\omega t + \phi)\right] . \end{aligned} \tag{5}$$

In the absence of any other forces $p = v_\mathrm{d}$ is constant and the oscillation center drifts at the constant velocity $q(t) = x(0) + v_\mathrm{d}t$. In particular, if $v_\mathrm{d} = 0$ (eg., $v_0 = 0$ and $\phi = 0$ or π), then $q(t) = x_0$ is constant. On adding a potential $V(x)$, we have in the transformed frame

$$H(q, p, t) = p^2/2 + V(q - \alpha[\cos(\omega t + \phi) - \cos\phi]) \ , \tag{6}$$

which can be conveniently expanded in a discrete Fourier series

$$H(q, p, t) = p^2/2 + \sum_{n=-\infty}^{\infty} V_n(q)\mathrm{e}^{in\omega t}, \tag{7}$$

as the perturbation is periodic with period $T = 2\pi/\omega$. The Fourier coefficient $V_0(q)$ is simply the time-averaged potential which is all that is required in the high-frequency approximation. Away from this limit, the other coefficients $V_n(q)$ contain valuable dynamical information. The ansatz that $V_n(q) \approx V_0(q)$ for a large number of n provides the other extreme. In this case, the Hamiltonian in the oscillating frame reduces to

$$\begin{aligned} H(q, p, t) &\approx p^2/2 + V_0(q) \sum_{n=-\infty}^{\infty} \mathrm{e}^{in\omega t}\,\mathrm{e}^{in\omega\phi} \\ &= p^2/2 + V_0(q) \sum_{n=-\infty}^{\infty} \delta(t - nT + \phi T/2\pi), \end{aligned} \tag{8}$$

using the Poisson sum rule. The phase ϕ can be set to zero with no loss of generality and integrating the classical equations of motion over one period T leads to the nonlinear,

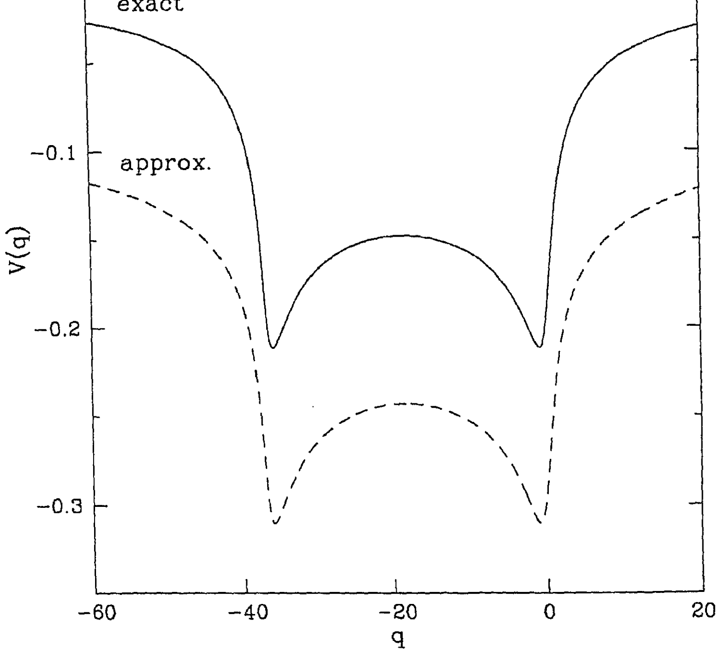


Figure 1. Kick potential as a function of q for $F = 5.0$ (a.u.) and $\omega = 0.52$ (a.u.), for both exact and approximate (evaluated only in the vicinity of the classical turning points) kicking terms. The curves clearly show that a simple constant shift is the only difference. Note that $F = 5$ means that the external field is five times larger than the binding field.

area-preserving map,

$$q_{n+1} = q_n + Tp_{n+1}, \quad p_{n+1} = p_n + G(q_n), \tag{9}$$

where q_n and p_n are the positions and momenta of the electron in the oscillating frame, evaluated once every period $T = 2\pi/\omega$ of the field. The total impulse of each kick $G(q_n)$ is the time-integral of the force over one period and can be simply expressed in terms of the space derivative of the time-integral of the oscillating potential,

$$G(q_n) = -\frac{\mathrm{d}}{\mathrm{d}q} TV_0(q)|_{q=q_n} \,. \tag{10}$$

A more physical motivation for the map as well as the explicit conditions for its validity are discussed in ref. [7]. Let us now consider the smoothed Coulomb potential

$$V(x) = -Z/\sqrt{a^2 + x^2}, \tag{11}$$

which in the moving frame is

$$V(q,t) = -Z/\sqrt{a^2 + (q - \alpha[\cos\omega t - 1])^2}, \tag{12}$$

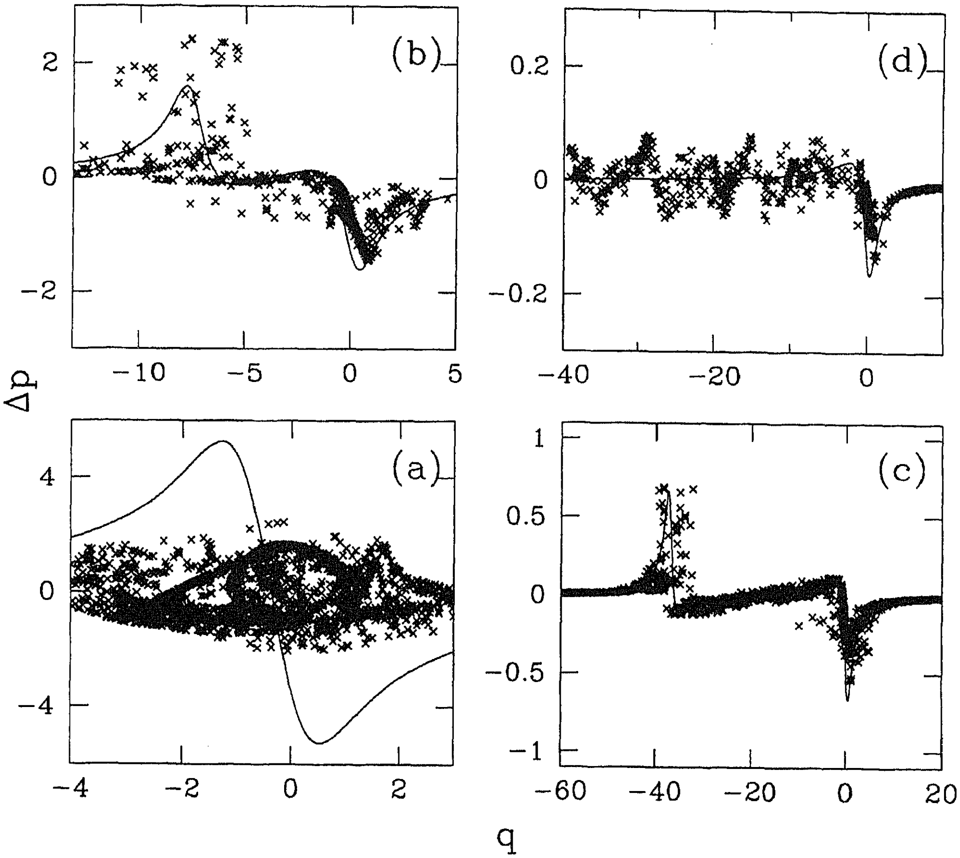


Figure 2. Comparison of the impulse generated over a single period from the differential dynamics (crosses) and the single valued function G for the map. The parameter α is **(a)** 0.37, **(b)** 3.7, **(c)** 18.5 and **(d)** 295.9. **(d)** displays only a small portion of the total range so as to emphasize the maximum discrepancy.

where Z is the nuclear charge and a eliminates the x^{-1} singularity at the origin and also defines a characteristic length scale. The time-average of $V(q,t)$ can be expressed exactly in terms of elliptic integrals though for our purposes it is enough to note that, as shown in figure 1, it is a double-welled potential with minima *near* the classical turning points at 0 and -2α for the free electron in the oscillating field. The map approximates the impulse only at these turning points and results in $G(q_n) = kF(q_n)$ where $k = \sqrt{8/F}$ and $F(x)$ involves complete elliptic integrals of the first and second kinds. The effective potential in the map approximation, also shown in figure 1, is simply shifted relative to the exact potential. The product $\kappa = kT$ is the *stochasticity parameter* and determines the nature (locally/globally/not chaotic) of the phase space of classical solutions. Note that both the limits of high field F and high frequency ω lead to small κ which corresponds to 'linear' or integrable dynamics.

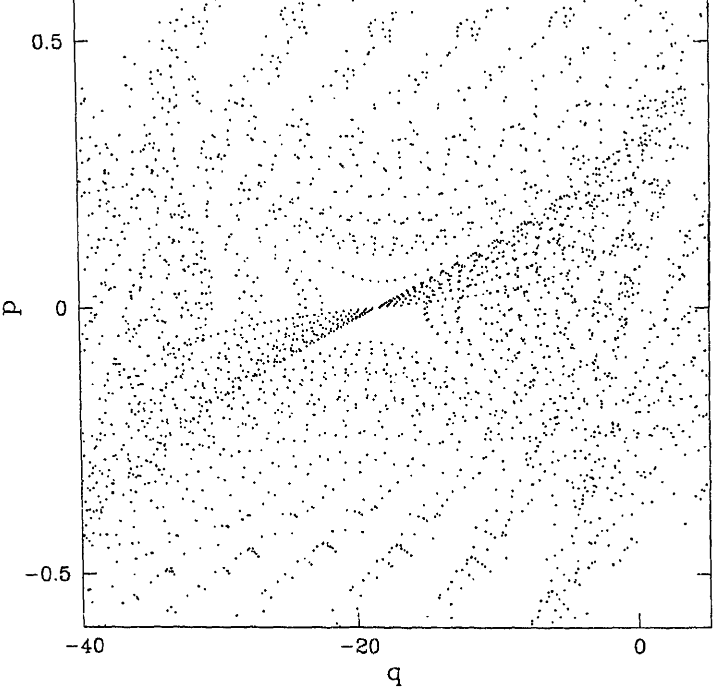


Figure 3. Poincaré section for $F = 5.0$ and $\omega = 0.52$ which is typical for large values of the stochasticity parameter. As discussed in the text, no stable (island) structures are expected and the classical dynamics is dominated by rapidly ionizing orbits.

The validity of the map approximation can be checked *a posteriori* by comparing the magnitude of the impulses experienced by the electron after each period of the oscillating field with those predicted by the single-valued function in the map. Figure 2 contrasts the impulse computed from the differential dynamics with G for a range of α. It is clear that the map becomes an effective description of the true dynamics with increasing α.

Having established the map as a good approximation to the true classical dynamics, we can now use a general stability analysis of trajectories to predict the classical parameter bounds for stabilization. This is also readily seen from the Poincaré surfaces of section which provide an overview of the classical dynamics. For example, figure 3 considers field parameters for which almost all the classical electron orbits are chaotic and ionizing. By contrast figure 4 shows a case where many initial conditions starting out near the nucleus are regular and do not ionize, tracing out regular island structures in the classical phase space. *Thus, classical prediction of stabilization is determined merely by the existence or otherwise of stable regions in phase space.*

The clearest representation of the parameter values for which regular regions exist in the classical phase space is a stability diagram in the space of $F - \omega$, shown in figure 5. The dotted line divides the overall parameter space into two regions where $\alpha < 1$ (on the right side) and $\alpha > 1$ (on the left side). When $\alpha \ll 1$ the map potential is

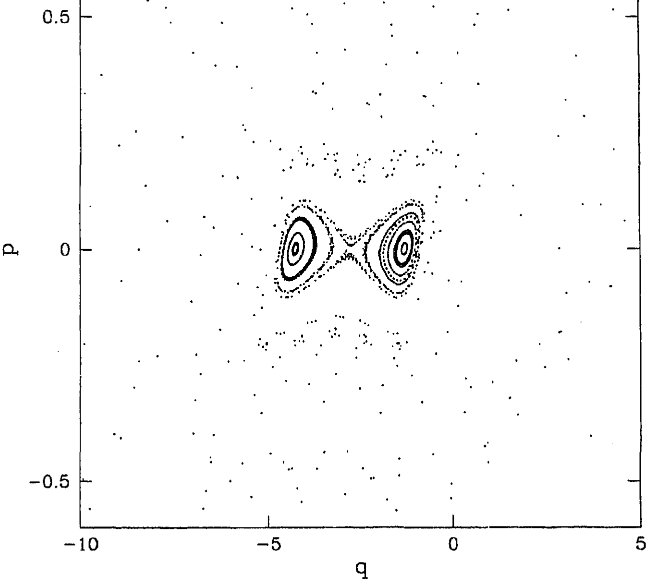


Figure 4. Poincaré section for parameters $\omega = 1.34, F = 5.0$ of showing the presence of stable classical islands. The occurrence of these stable regions is typical for smaller values of the stochasticity parameter.

a single well centered on a fixed point near $-\alpha$. Increasing α leads to the effective double-well potential shown earlier with two elliptic (stable) and one hyperbolic (unstable) fixed points. The elliptic fixed points become unstable for parameter values below an analytically predicted stability border (solid line) after which *no classical mechanism for stabilization exists.* The predicted boundary is consistent with the existence of stable structures in both map dynamics (indicated by the dotted region) and the full differential dynamics (indicated by the open circles), as seen from figure 5. Note the region of low F and ω where map and differential dynamics disagree. In this regime, the relevant time scale is the internal period rather than the external field period we use. This results in a different stroboscopic approximation called the Kepler map [12].

We consider three parameter sets corresponding to the points A, B and C in the stability diagram. Purely classical stability arguments would suggest no stabilization for A and a larger fraction for C as compared with B. C is also closer to the line $\alpha = 1$ which means a near single-well effective potential. As I shall now show, the quantum determination of stability is considerably more complicated. For the quantum dynamics, we simply integrate the time-dependent Schrödinger equation (in atomic units)

$$i\,\frac{\partial\Psi(x,t)}{\partial t} = \left[-\frac{1}{2}\frac{\partial^2}{\partial x^2} - V(x) - xE(t)\right]\Psi(x,t), \tag{13}$$

on a space-time grid using the standard two-sweep algorithm [13]. The initial condition is the ground state of the undriven potential. Given our picture of the stabilization

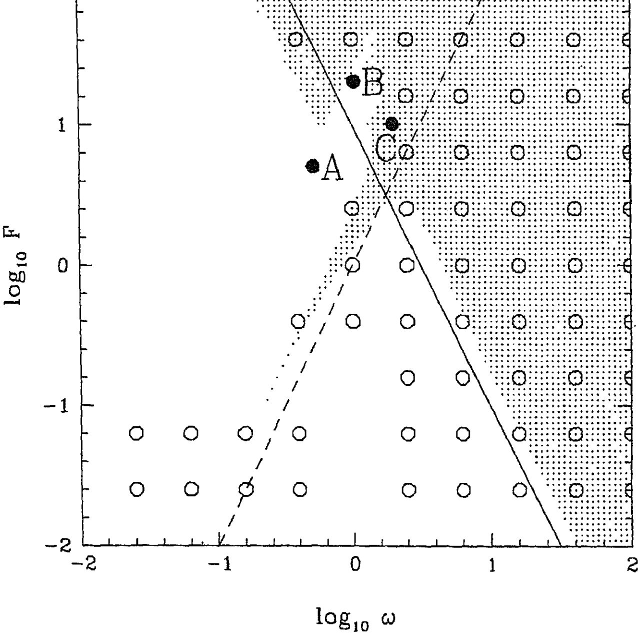


Figure 5. The classical stability of the map (dots) and the full differential dynamics (circles) was assessed by advancing the equations of motion 200 periods of the perturbation. If the trajectory returned to the vicinity of the nucleus a point was plotted in the $(\omega - F)$ parameter plane. The dividing line $\alpha = 1.0$ is indicated by the dotted line while the stability boundary is shown by the solid line.

mechanism as resulting from a free-particle interacting periodically with trapping centers (in this case a double well potential), a simple measure of ionization is the fraction of wavefunction that is outside the "interaction volume". A smaller ionized fraction means increased stabilization. As the maximum spacing of the double-wells (averaging over an external period) is 4α, a suitable choice for this interaction volume (in one dimension) is $-4\alpha \leq x \leq 4\alpha$. Thus, a definition of ionization as

$$P_{\mathrm{I}}(t) = \int_{|x|>4\alpha} |\Psi(x,t)|^2 \mathrm{d}x \tag{14}$$

is adequate to establish when stabilization is significant.

The results of the quantum simulations for cases A, B and C are shown in the lower two panels in figure 6. The corresponding classical phase portraits shown reinforce our inferences from the stability diagram; no stabilization for A while larger islands exist for C as compared with B. However, the ionized fraction as calculated from the quantum evolution supports the contrary result that there is *more stabilization for* A *as compared with* B. Case C is the most stable which is at least consistent with the classical prediction. What is the origin of this discrepancy?.

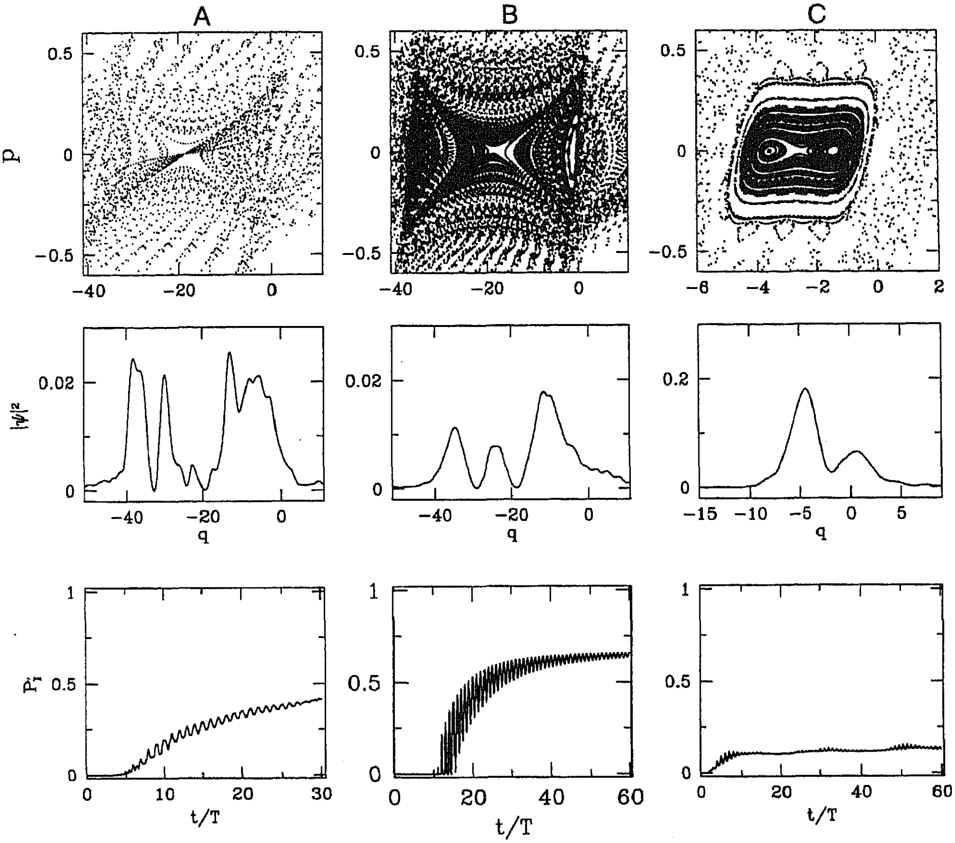


Figure 6. Classical phase portraits (upper panel), residual quantum wavefunctions (middle panel), and ionization probability versus time (in units of the period T) (bottom panel). The parameters are (**A**) $F = 5.0$, $\omega = 0.52$; (**B**) $F = 20$, $\omega = 1.04$; and (**C**) $F = 10$ and $\omega = 2.0$. Note that the peak structure of the final wavefunction reflects both stable and unstable classical fixed points. For case C, the peaks are beginning to coalesce reflecting the approach of the single-well effective potential (see text).

The resolution comes from the fact that the classical predictions were based entirely on the stability analysis of low-order fixed points. Though a fixed point may be unstable, there remains an important dynamically invariant, but still unstable, classical structure – the *homoclinic tangle*. Homoclinic tangles corresponding to cases A and B are shown in figures 7 and 8 respectively. The tangle emanates from unstable fixed points and its complicated appearance is simply a consequence of the area-preserving constraint of Hamiltonian evolution. The unstable tangle also provides support for quantum wavefunctions – an example of a phenomenon referred to as ‘scarring’ [14]. A heuristic estimate of the ‘support’ provided by the tangle for quantization is obtained by considering how the size (area) of the tangle compares with Planck's constant. This measure clearly shows that despite larger stable regions, the overall support *stable*

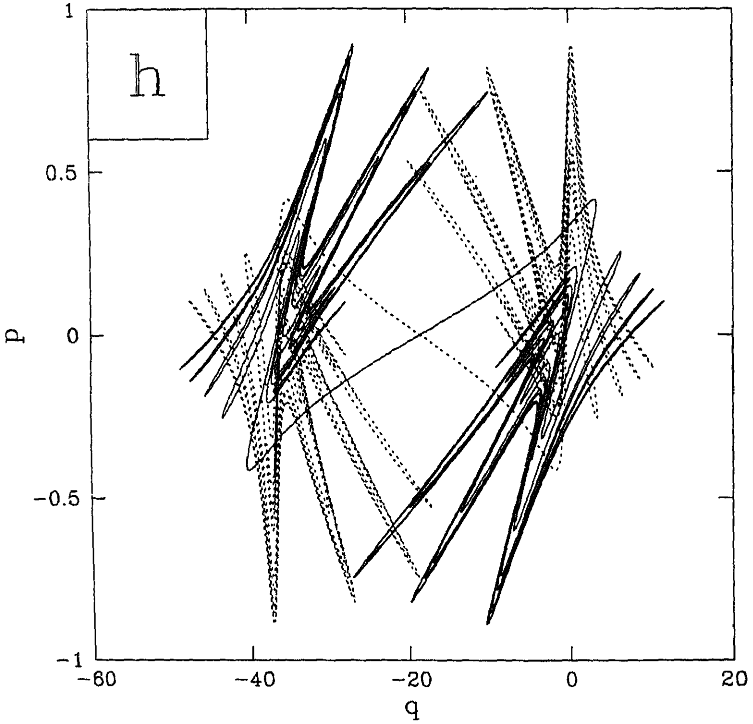


Figure 7. Homoclinic tangle associated with the fixed point at $(-\alpha, 0)$ for case A. Near the fixed point, the solid line gives the unstable direction while the dashed line is the stable direction. The size of Planck's constant h is shown to illustrate that several states can be supported by the single structure. An estimate of the number of states is given by the number of h boxes needed to cover the structure.

and unstable, is considerably less in case B as compared with case A. This is the reason for the reduced stabilization. However, the wavefunction at the end of the interaction time in case B exhibits three distinct peaks which clearly reflect the stable regions in the phase space, unlike A. Case C is one where the phase space is dominated by a large stable region and classical and quantum intuition agree. Thus, the discrepancy in cases A/B is a direct consequence of what was stated to be quantum chaology – "..nonclassical behaviour characteristic of systems whose classical motion exhibits chaos".

These considerations can be extended to the full 3-D system where a four-dimensional map and considerably more involved analysis is necessary [15, 16], though scarring is once again a relevant feature. For those concerned by the seemingly unphysically large field strengths used in the illustrative cases, it should be noted that these results can be scaled to other frequency regimes and also to excited state initial conditions, where an experimental realization is more likely [17]. The mobility of the light component of the two-component plasmas considered in wakefield accelerators is another potential experimental system.

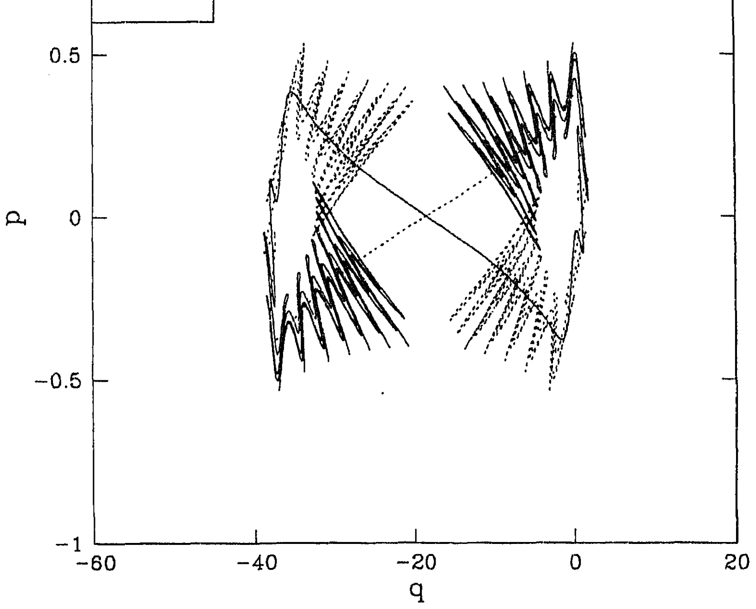


Figure 8. Same as figure 7 but for case B. Note that the classical excursion $\alpha = F/\omega^2$ is the same in the two cases.

3. Atom optics and dynamical localization

Our earlier discussion makes clear that quantum dynamics of systems with chaotic classical limit are usually much more stable in their evolution than their classical counterparts. However, contrary to expectations, the mechanisms that lead to quantum stabilization are not solely associated with classically stabilizing phase space structures like the regions around elliptic periodic orbits, KAM-tori, and separatices. Even structures connected with unstable classical motion such as cantori, hyperbolic periodic orbits, homoclinic and heteroclinic tangles can support quantum mechanical wavefunctions. Even though these ‘scarred’ wavefunctions are ubiquitous in both autonomous and non-autonomous strongly coupled quantum systems, they form a small part of the bigger picture of quantum dynamics with chaotic classical limit.

For example there is also a genuinely quantum dynamical stabilization mechanism. ‘Dynamical localization’ refers to the fact that eigenstates of quantum dynamics can be localized in an action variable like momentum despite the fact that the classical limit exhibits (deterministic) diffusion [18]. This effect is analogous to Anderson localization in tight-binding models [19] and provides a global mechanism (does not distinguish details of the classical phase space) for the quantum suppression of classical chaos. In one-dimension, the localized wavefunction has a characteristic exponential form as in the case of Anderson localization.

This is in contrast with the phenomenon of 'scarring' where the quantum wavefunctions show enhanced probability around specific classically unstable invariant structures. Thus, the excitation of scarred wavefunctions provides a 'local' mechanism of suppression of the classically chaotic dynamics.

As the emphasis in this article is on the usefulness of these ideas in unexpected contexts, I will now consider a recent example of dynamical localization in atom optics. Instead of the smoothed Coulomb potential, let us take $V(q) = k\cos q$. In the presence of the time-drive and in the moving frame, we have

$$\begin{aligned} H &= p^2/2 + V(q - \alpha\sin(\omega t)) \\ &= p^2/2 + k\cos(q - \alpha\sin(\omega t)) \\ &= p^2/2 + k\sum_n J_n(\alpha)\cos(q - n\omega t), \end{aligned} \tag{15}$$

which describes a modulated pendulum where the Fourier weights J_n are Bessel functions of order n. For large α, the asymptotic form for J_n can be taken where the amplitude varies slowly with n (as required by the ansatz of the earlier example). The resulting map is a textbook paradigm for Hamiltonian dynamics and chaos – the standard or Chirikov–Taylor map. More significantly, the quantum standard map or kicked rotor is *the* paradigm for dynamical localization.

A novel realization of this driven system was reported recently by considering the momentum transfer from a modulated standing wave of light to a sample of independent, ultra-cold atoms [3, 20]. The thermal (cold) sample of approximately 10^5 sodium atoms provide a Gaussian initial condition in momentum. The relevance of the internal structure of the atom to the dynamics can be eliminated by tuning the frequency of the standing wave far from resonance. The remaining dynamics is that of a 'structureless' atom, described by the above Hamiltonian, which undergoes a momentum change of $2\hbar k_L$ (k_L is the wavenumber of the light wave) when the velocities of the atom and standing wave are matched. These 'resonant kicks' occur twice every standing wave period though they are not equally spaced in time. The experiment measures the final momentum distribution of the atoms as the modulation amplitude of the standing wave is varied.

The classical dynamics generated by (15) is governed by resonances at $p = \dot{q} = n\omega$ with approximate widths $4\sqrt{k|J_m(\alpha)|}$ in momentum. These resonances overlap as α is increased after which the classical particle can diffuse in momentum. It should be noted that there is a maximum momentum value $p_{\max} = \alpha$ beyond which the interaction turns off. This is easily seen as the velocities of the atom and standing wave have to match (stationary phase condition) for the momentum transfer to occur. Thus, the classical diffusion is over a bounded region in momentum space, the extent of which can be controlled by changing α.

The first column shown in Figure 9 corresponds to zero modulation amplitude, which is just the case of the simple pendulum. The classical, experimentally measured and quantum distribution of final momenta of the atomic ensemble agree once again because the phase space is entirely regular. On increasing α, the phase space becomes increasingly chaotic and the stable regions shrink as shown in column two. The dominant classical mechanism is now diffusion in momentum. The

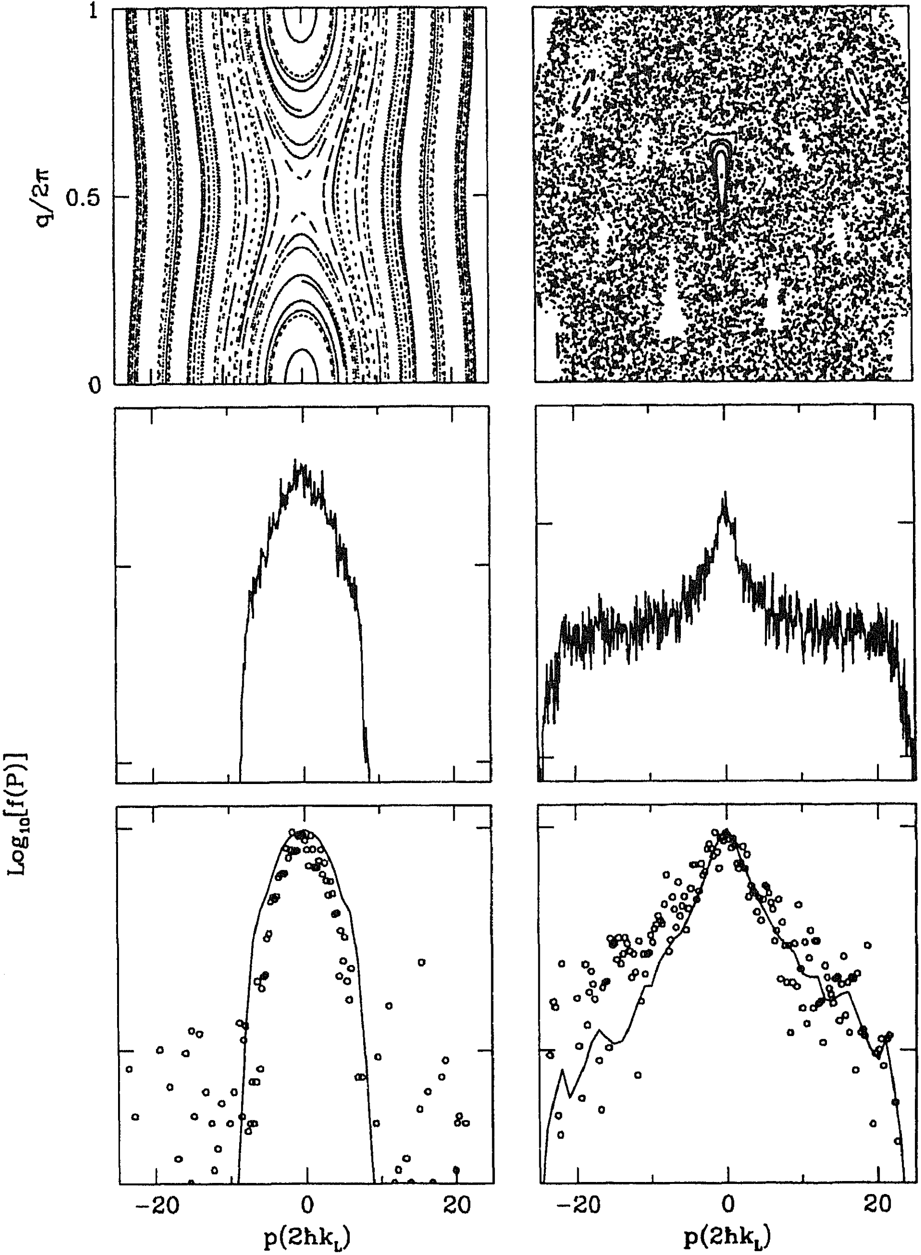


Figure 9. Phase portraits (upper panel), classical momentum distributions (middle panel), and experimentally measured momentum distributions compared with quantum theory (bottom panel, theory marked by lines). The first column is for $\alpha = 0$ and is just the case of the simple pendulum while $\alpha \neq 0$ in column two. The initial condition corresponds to a Gaussian distribution in momentum and fully extended (plane waves) in position. The vertical scales for the distributions are logarithmic and are marked in decades.

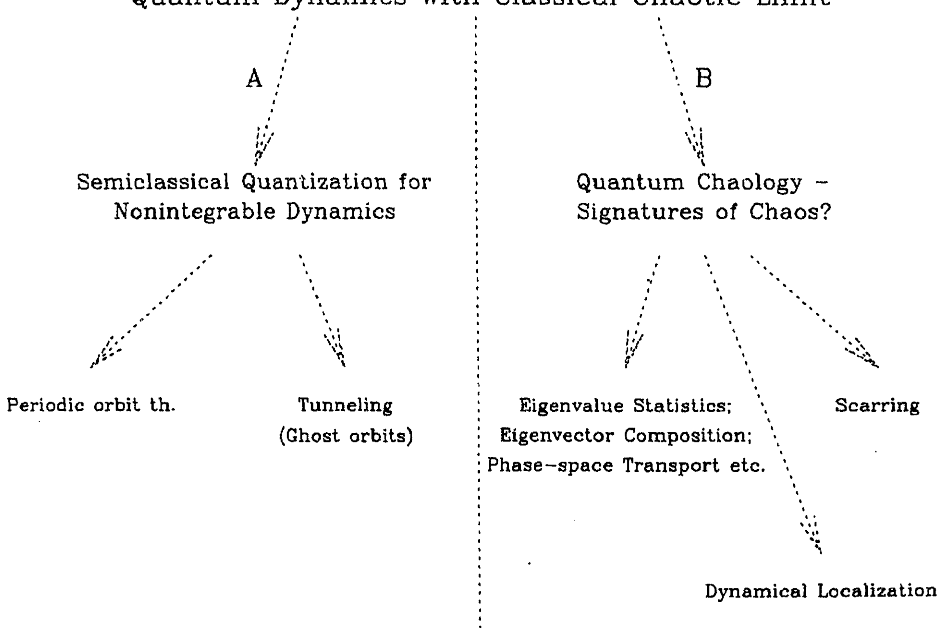

Quantum Dynamics with Classical Chaotic Limit
A
B
Semiclassical Quantization for Nonintegrable Dynamics
Quantum Chaology – Signatures of Chaos?
Periodic orbit th.
Tunneling (Ghost orbits)
Eigenvalue Statistics; Eigenvector Composition; Phase–space Transport etc.
Scarring
Dynamical Localization

Figure 10. Schematic representation of issues which arise when considering the quantization of classically chaotic dynamics.

final classical momentum distribution is nearly uniform over the allowed region, except for a small peak corresponding to the remnant stable region. This is in sharp contrast with the quantum and experimental distributions which both exhibit an exponential distribution in momentum (note the logarithmic *y*-axis). This is an example of dynamical localization.

Anderson localization describes the motion of electrons on a disordered lattice where the randomness is externally imposed. In dynamical localization randomness or, more correctly, pseudorandomness is generated by the dynamics leading to localization in an action variable, rather than in space. Dynamical equivalents of higher dimensional Anderson models [21] also exist which should be realized soon by the atom optics experiments. It should be mentioned that these experiments have already confirmed many of the details of the theory of dynamical localization [22] as well as the mechanism of overlapping resonances to achieve global chaos [23].

4. The bigger picture

Scarring and dynamical localization are merely two aspects of a much larger view of quantum dynamics with classical chaotic limit. As shown by the schematic in figure 10, this picture originates from two simple questions which still demarcate current approaches. (a) How does semiclassical quantization work for nonintegrable dynamics? and (b) What are the semiclassical signatures of chaos?

The first question resulted in periodic orbit quantization, an idea which dates back to Van Vleck in 1928 and which was extensively pursued by Gutzwiller [24]. The seminal idea is the use of classical trajectories to calculate the quantum propagator of a mechanical system within the semiclassical approximation. The trace of this propagator, which gives information about the spectrum of the dynamics, is therefore linked to the periodic orbits of the classical dynamics. This approach works not only for integrable dynamics but also for strongly chaotic dynamics as well. It has been applied successfully to describe the quantum properties of model systems like the anisotropic Kepler problem [25, 26] and the motion of a particle on a surface of constant negative curvature, where the trace-formula is exact (see for example [27]); as well as to the spectral analysis of a hydrogen atom in a strong magnetic field [8] and the doubly excited spectrum of the helium atom [10]. A related and yet outstanding problem is periodic orbit quantization for dynamics in a mixed phase space, where regular and chaotic structures coexist.

A recent twist to the periodic-orbit analysis came with the observation that in systems where new orbits are created through bifurcation, the impact of these orbits is felt in the quantum spectrum *before they appear as real orbits of the classical dynamics* (parametrically speaking) [28, 29, 30]. These so-called "ghost" orbit contributions further modify the semi-classical analysis.

The second question leads to "quantum chaology". It includes the study of the statistical features of the eigenvalue spectra, as well as of the eigenvector components in a fixed representation. Using this approach, a variety of interesting results have surfaced (see for example [31, 32]). An example of consequence to experiments is the relationship between the statistical features of the spectra of most strongly coupled systems to certain matrix ensembles devised to understand nuclear spectra [33]. However, this statistical description results in randomized eigenfunctions and has to be modified as soon as a dynamically localized or scarred wavefunction is observed.

The interplay of the different mechanisms had not been addressed till very recently. In large part this is due to the fact that the demands made on model systems by the two branches are quite distinct. For example, most aspects of quantum chaology can be studied within the δ-kicked rotor or standard map paradigm. However, semiclassical or periodic orbit quantization is greatly facilitated by the existence of a symbolic coding (alphabet) and pruning rules (grammar) for the classical periodic orbits, as in the 3- and 4-disk scattering systems. It is not easy to calculate classical invariant structures in the standard map. As such even the competition between local (scarring) and global (dynamical localization) suppression mechanisms has not been explored. Scarring [34] and dynamical localization [35] have only recently been reconciled with semiclassical theory, though even there open questions still remain.

Piecewise continuous potentials have been used to address a number of issues related to classical transport [36]. Scharf and I [37] showed that a piecewise linear standard map [38] facilitated the calculation of classical invariant structures for scarring and provided a simple coding for periodic orbits. We used this modified dynamics to explore the strength of scarring [39] as well as the construction of local effective Hamiltonian flows based on classical periodic orbits [40]. This local effective description becomes necessary in mixed phase spaces though the crucial issue of matching solutions then becomes the open question. Current work on relating these local effective flows to Frobenius–Perron [41] operator dynamics appears very promising.

Experimentally, a strong focus of the next generation of experiments in atom optics will be mixed phase space dynamics and the effects of noise [42]. Direct excitation of scarred states and an assessment of their coherence with an atom interferometer may also be possible. Advances in fabrication technology have opened up another new testing ground in the area of layered heterostructures. The role of the ideas we have discussed to the transport or magnetic properties of driven quantum wells and a variety of mesoscopic systems is currently under intense scrutiny with the anticipation that these might impact on device physics as well. In particular, experiments measuring changes in the magnetoconductance as the shape of the scattering domain, for a two-dimensional electron gas, is changed from integrable (circle or square) to nonintegrable (stadium) [43, 44] provide interesting deviations from the theoretical paradigms. We had suggested that ghost orbit contributions are important to the dynamics once the Lorentz force is considered in computing classical trajectories [30]. Recently experimental evidence also indicates deviations from random matrix theory predictions [45] as well as the role of scarring [46]. Inelastic scattering and decoherence are also issues which invoke the role of parametric or systemic noise on any stabilized quantum structure [47]. It is my view that just as turbulence provides strong motivation in classical dynamical analyses, these inherently ‘messy’, truly many-body systems pose an analogous challenge for nonintegrable quantum dynamical systems.

Acknowledgements

The author would like to acknowledge fruitful collaborations with R V Jensen and the experimental group of M G Raizen. The ‘big picture’ developed over many years of working with his friend and ‘sounding board’ for ideas, Rainer Scharf. This work was supported in part by the U.S. National Science Foundation.

References

[1] M V Berry, *Proc. R. Soc. London* **A423**, 219 (1989)
[2] See, for example, *Quantum chaos: between order and disorder: a selection of papers* edited by G Casati and B V Chirirov (Cambridge University Press, New York, 1995)
[3] F L Moore, J C Robinson, C F Bharucha, P E Williams and M G Raizen, *Phys. Rev. Lett.* **73**, 2974 (1994)
[4] P W Milonni and B Sundaram, *Progress in optics* **XXXI**, 3 (1993) for a review of intense-field physics
[5] Q Su, J H Eberly and J Javanainen, *Phys. Rev. Lett.* **64**, 862 (1990)
[6] K C Kulander, K J Schafer and J L Krause, in *Multiphoton processes* edited by G Mainfray and P Agostini (CEA-Press, Paris, 1991); *Phys. Rev. Lett.* **66**, 2601 (1991)
[7] R V Jensen and B Sundaram, *Phys. Rev. Lett.* **65**, 1964 (1990); *Phys. Rev.* **A47**, 1415 (1993)
[8] D Wintgen, *Phys. Rev. Lett.* **61**, 1803 (1988)
H Friedrich and D Wintgen, *Phys. Rep.* **183**, 37 (1989)
D Delande in *Chaos and quantum physics* edited by M-J Giannoni, A Voros and J Zinn-Justin (Elsevier, London, 1991)
[9] R V Jensen, M M Sanders, M Saraceno and B Sundaram, *Phys. Rev. Lett.* **63**, 2771 (1989)
[10] D Wintgen, K Richter and G Tanner, *Chaos* **2**, 19 (1992)
G Ezra, K Richter, G Tanner and D Wintgen, *J. Phys.* **B24**, L413 (1991)

[11] This is true only with respect to the unperturbed levels. The validity of semiclassical arguments becomes clear on considering the distorted levels, which bear no resemblance to the unperturbed levels under the extremely strong perturbation

[12] G Casati, I Guarneri and D L Shepelyansky, *IEEE J. Quantum Electron.* **24**, 1420 (1988)

[13] S E Koonin and D C Meredith, *Computational physics* (Addison Wesley, Menlo Park, California, 1990)

[14] E J Heller, *Phys. Rev. Lett.* **53**, 1515 (1984)
E J Heller and S Tomsovic, *Phys. Today*, July 1993, p. 38 and references therein

[15] R V Jensen and B Sundaram, *Phys. Rev.* **A47**, 1415 (1993); *Laser Phys.* **3**, 291 (1993)

[16] F Benvenuto, G Casati and D L Shepelyansky, *Z. Phys.* **B94**, 481 (1994)

[17] R R Jones and P H Bucksbaum, *Phys. Rev. Lett.* **67**, 3215 (1991)
M P de Boer, J H Hoogenraad, R B Vrijen, L D Noordam and H G Muller, *Phys. Rev. Lett.* **71**, 3263 (1993)

[18] F M Izrailev, *Phys. Rep.* **196**, 299 (1990)

[19] S Fishman, D R Grempel and R E Prange, *Phys. Rev. Lett.* **49**, 509 (1982); *Phys. Rev.* **A29**, 1639 (1984)

[20] J C Robinson, C F Bharucha, F L Moore, R Jahnke, G A Georgakis, Q Niu, M G Raizen and B Sundaram, *Phys. Rev. Lett.* **74**, 2974 (1995)

[21] G Casati, I Guarneri, F Izrailev and R Scharf, *Phys. Rev. Lett.* **64**, 5 (1990)

[22] F L Moore, J C Robinson, C F Bharucha, B Sundaram and M G Raizen, *Phys. Rev. Lett.* **75**, 4598 (1995)

[23] J C Robinson, C F Bharucha, K Madison, B Sundaram, S R Wilkinson, F L Moore and M G Raizen, *Phys. Rev. Lett.* **76**, 3304 (1996)

[24] M C Gutzwiller, *Chaos in classical and quantum mechanics* (Springer, Berlin, 1990)

[25] M C Gutzwiller, *Phys. Rev. Lett.* **45**, 150 (1980)

[26] M C Gutzwiller, *Physica* **D5**, 183 (1982)

[27] M Sieber and F Steiner, *Phys. Lett.* **A144**, 159 (1990)

[28] M Kuś, F Haake and D Delande, *Phys. Rev. Lett.* **71**, 2167 (1993)

[29] R Scharf and B Sundaram, *Phys. Rev.* **E49**, R4767 (1994)

[30] B Sundaram and R Scharf, *Physica* **D83**, 257 (1995)

[31] B Eckhardt, *Phys. Rep.* **163**, 205 (1988)

[32] F Haake, *Quantum signatures of chaos* (Springer, Berlin, 1991)

[33] T A Brody, J Flores, J B French, P A Mello, A Pandey and S S M Wong, *Rev. Mod. Phys.* **53**, 385 (1981)

[34] O Agam and S Fishman, *J. Phys.* **A26**, 2113 (1993)
O Agam and N Brenner, *J. Phys.* **A28**, 1345 (1995)

[35] R Scharf and B Sundaram, *Phys. Rev. Lett.* **76**, 4907 (1996)

[36] See, for example, I Dana, *Phys. Rev. Lett.* **64**, 2339 (1990)
T Tél, *J. Stat. Phys.* **33**, 195 (1983) and references therein

[37] R Scharf and B Sundaram, *Phys. Rev.* **A43**, 3183 (1991)

[38] See J Wilbrink, *Physica* **D26**, 358 (1986)
S Bullet, *Commun. Math. Phys.* **107**, 241 (1986)
Koo-Chul Lee, *Physica* **D35**, 186 (1989) for discussions of differences between the classical standard and piecewise linear maps

[39] R Scharf and B Sundaram, *Phys. Rev.* **A45**, 3615 (1992)

[40] R Scharf and B Sundaram, *Phys. Rev.* **A46**, 3164 (1992)

[41] H H Hasegawa and D J Driebe, *Phys. Rev.* **E50**, 1781 (1994)

[42] M G Raizen, personal communication

[43] Readers are directed to *Chaos* **3**, 417–691 (1993) for a collection of articles on chaotic scattering, both on theory and experiments

[44] C M Marcus, A J Rimberg, R M Westervelt, P F Hopkins and A C Gossard, *Phys. Rev. Lett.* **69**, 506 (1992)
C M Marcus, R M Westervelt, P F Hopkins and A C Gossard, *Chaos* **3**, 643 (1993)

[45] A M Chang, H U Baranger, L N Pfeiffer, K M West and T Y Chang, *Phys. Rev. Lett.* **76**, 1695 (1996)

J A Folk, S R Patel, S F Godijn, A G Huibers, S M Cronenwett, C M Marcus, K Campman and A C Gossard, *Phys. Rev. Lett.* **76**, 1699 (1996)

[46] P B Wilkinson, T M Fromhold, L Eaves, F W Sheard, N Muira and T Takamasu, *Nature* **380**, 608 (1996)

[47] See R Scharf and B Sundaram, *Phys. Rev.* **E49**, R2509 (1994)
L Sirko, M R W Bellermann, A Haffmans, P M Koch and D Richards, *Phys. Rev. Lett.* **71**, 2895 (1993) for theoretical and experimental examples, respectively

physics
pp. 487–501

Periodic orbits in polygonal billiards

DEBABRATA BISWAS
Center for Chaos and Turbulence Studies, Niels Bohr Institute Blegdamsvej 17, Copenhagen Ø, Denmark
On leave from: Theoretical Physics Division, Bhabha Atomic Research Centre, Bombay 400085, India

Abstract. We review some properties of periodic orbit families in polygonal billiards and discuss in particular a sum rule that they obey. In addition, we provide algorithms to determine periodic orbit families and present numerical results that shed new light on the proliferation law and its variation with the genus of the invariant surface. Finally, we deal with correlations in the length spectrum and find that long orbits display Poisson fluctuations.

Keywords. Polygonal billiards; periodic orbits; proliferation law; statistics.

PACS Nos 05.45; 3.65

1. Introduction

Billiards are an interesting and well studied class of hamiltonian systems that display a wide variety of dynamical behaviour depending on the shape of the boundary. The ray equations commonly considered follow from a short wavelength expansion of the Schrödinger equation with Dirichlet (or Neumann) boundary conditions. A particle thus moves freely between collisions at the boundary where it suffers specular reflection. As an example, a square billiard generates regular dynamics that is restricted to a torus in phase space due to the existence of two well behaved constants of motion. In contrast, generic trajectories in the enclosure formed by three intersecting discs explore the entire constant energy surface and hence the system is ergodic. Moreover, orbits that are nearby initially move exponentially apart with time and hence the system is said to be metrically chaotic.

Polygonal billiards are a class of systems whose properties are not as well known. The ones accessible to computations are rational angled polygons and they generically belong to the class of systems referred to as pseudo-integrable [1]. They possess two constants of motion like their integrable counterparts [2] but the invariant surface Γ is topologically equivalent to a sphere with multiple holes [1] and not a torus. Such systems are characterized by the genus g where $2g$ is the numbers of cuts required in Γ to produce a singly connected region (alternately, g is the number of holes in Γ). As an example, consider the 1-step billiard in figure 1. For any trajectory, p_x^2 and p_y^2 are conserved. The invariant surface consists of four sheets (copies) corresponding to the four possible momenta, $(\pm p_x, \pm p_y)$ that it can have and the edges of these sheets can be identified such that the resulting surface has the topology of a double torus.

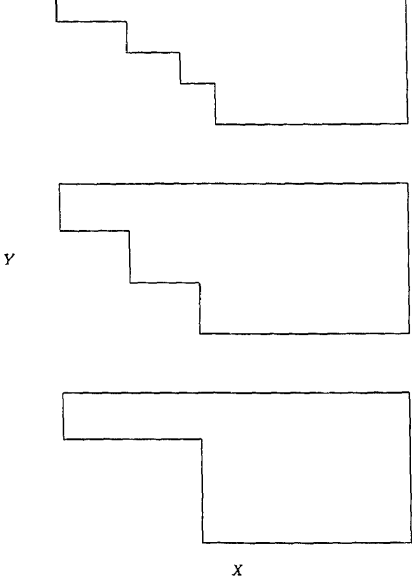


Figure 1. The 1, 2 and 3-step billiards are pseudo-integrable due to the $3\pi/2$ internal angles. The 1-step billiard (bottom) is also referred to as the L-shaped billiard in the text.

The classical dynamics no longer has the simplicity of an integrable system where a transformation to *action* and *angle* co-ordinates enables one to solve the global evolution equations on a torus. On the other hand, the dynamics is non-chaotic with the only interesting feature occurring at the singular vertex with internal angle $3\pi/2$. Here, families of parallel rays split and traverse different paths, a fact that limits the extent of periodic orbit families. This is in contrast to integrable billiards (and to the $\pi/2$ internal angles in figure 1) where families of rays do not see the vertex and continue smoothly.

We shall focus here on the periodic orbits of such systems for they form the skeleton on which generic classical motion is built. They are also the central objects of modern semiclassical theories [3] which provide a duality between the quantum spectrum and the classical length and stabilities of periodic orbits. In polygonal billiards, primitive periodic orbits can typically be classified under two categories depending on whether they suffer even or odd number of bounces at the boundary. In both cases however, the linearized flow is marginally unstable since the Jacobian matrix, J_p, connecting the transverse components ($u_\perp(t+T_p) = J_p\, u_\perp(t)$ where $u_\perp = (q_\perp, p_\perp)^T$ and T_p is the time period of the orbit) has unit eigenvalues. However, when the number of reflections, n_p, at the boundary is even, the orbit occurs in a 1-parameter family while it is isolated when n_p is odd. This follows from the fact that the left (right) neighbourhood of an orbit becomes the right (left) neighbourhood on reflection so that an initial neighbourhood can never overlap with itself after odd number of reflections [1].

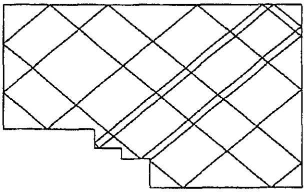


Figure 2. An example of a periodic orbit in the 3-step billiard. Note that adjacent parallel rays are also periodic so that a family of periodic orbits exists. Its extent is limited by the $3\pi/2$ vertices.

While isolated orbits are important and need to be incorporated in any complete *periodic orbit theory*, a family of identical orbits has greater weight and such families proliferate faster than isolated orbits [4]. Besides, in a number of cases including the L-shaped billiard of figure 1, isolated periodic orbits are altogether absent since the initial and final momenta coincide only if the orbit undergoes even number of reflections. For this reason, we shall consider families of periodic orbits in this article (see figure 2 for an example).

Not much is however known about the manner in which they are organized and the few mathematical results that exist [4] concern the asymptotic properties of their proliferation rate. For a sub-class of rational polygons where the vertices and edges lie on an integrable lattice (the so called almost-integrable systems [4]), these asymptotic results are exact. It is known for example that the number of periodic orbit families, $N(l)$, increases quadratically with length, l, as $l \rightarrow \infty$. For general rational polygons, rigorous results show that $c_1 l^2 \leq N(l) \leq c_2 l^2$ for sufficiently large values of l [4, 5] while in case of a regular polygon P_n with n sides, it is known that $N(l) \simeq c_n l^2/A$ [6] where c_n is a number theoretic constant and A denotes the area of P_n. Very little is known however about other aspects such as the sum rules obeyed by periodic orbit families in contrast to the limiting cases of integrable and chaotic behaviour where these have been well studied. Besides, it is desirable to learn about the variation of the proliferation rate as a function of the genus for this should tell us about the transition to chaos in polygonal approximations of chaotic billiards [7, 8].

We shall concern ourselves primarily with a basic sum rule obeyed by periodic orbits arising from the conservation of probability. This leads to the proliferation law, $N(l) = \pi b_0 l^2/\langle a(l)\rangle$ where b_0 is a constant. The quantity $\langle a(l)\rangle$ is the average area occupied by all families of periodic orbits with length less than l and is not a constant unlike integrable billiards. We provide here the some numerical results on how $\langle a(l)\rangle$ changes with the length of the orbits and the genus of the invariant surface. While this does not allow us to make quantitative predictions, the qualitative behaviour sheds new light on the proliferation law for short orbits and its variation with the genus, g of the invariant surface.

Finally, we shall also study correlations in the length spectrum of periodic orbits. The numerical results provided here are the first of its kind for generic pseudo-integrable billiards and corroborate theoretical predictions provided earlier [9].

The organization of this paper is as follows. In § 2, we provide the basic sum rule and the proliferation law in pseudo-integrable systems. In § 3, we discuss algorithms to determine periodic orbit families in generic situations. This is followed by a numerical are summarized in §6.

2. The basic sum rule

The manner in which periodic orbits organize themselves in closed systems is strongly linked to the existence of sum rules arising from conservation laws. For example, the fact that a particle never escapes implies that for chaotic systems [10, 11]

$$\left\langle \sum_p \sum_{r=1}^{\infty} \frac{T_p \delta(t - rT_p)}{|\det(\mathbf{1} - \mathbf{J}_p{}^r)|} \right\rangle = 1 \tag{1}$$

where the summation over p refers to all primitive periodic orbits, T_p is time period, $\mathbf{J}_p$ is the stability matrix evaluated on the orbit and the symbol $\langle\cdot\rangle$ denotes the average value of the expression on the left. Since the periodic orbits are unstable and isolated, $|\det(\mathbf{1} - \mathbf{J}_p{}^r)| \simeq e^{\lambda_p r T_p}$, where λ_p is the Lyapunov exponent of the orbit. The exponential proliferation of orbits is thus implicit in (1).

A transparent derivation of (1) follows from the classical evolution operator

$$L^t \circ \phi(\mathbf{x}) = \int \mathrm{d}\mathbf{y} \delta(\mathbf{x} - \mathbf{f}^t(\mathbf{y}))\phi(\mathbf{y}) \tag{2}$$

where L^t governs the evolution of densities $\phi(\mathbf{x})$, $\mathbf{x} = (\mathbf{q}, \mathbf{p})$ and $\mathbf{f}^t$ refers to the flow in the full phase space. We denote by $\Lambda_n(t)$ the eigenvalue corresponding to an eigenfunction $\phi_n(\mathbf{x})$ such that $L^t \circ \phi_n(\mathbf{x}) = \Lambda_n(t)\phi_n(\mathbf{x})$. The semi-group property, $L^{t_1} \circ L^{t_2} = L^{t_1 + t_2}$, for continuous time implies that the eigenvalues $\{\Lambda_n(t)\}$ are of the form $\{e^{\lambda_n t}\}$. Further, for hamiltonian flows, (2) implies that there exists a unit eigenvalue corresponding to a uniform density so that $\lambda_0 = 0$. For strongly hyperbolic systems, $\lambda_n = -\alpha_n + i\beta_n, n > 1$ with a negative real part implying that

$$\mathrm{Tr}\, L^t = 1 + \sum_n \exp\{-\alpha_n t + i\beta_n t\}. \tag{3}$$

Equation (1) is thus a restatement of (3) with the trace expressed in terms of periodic orbit stabilities and time periods.

For polygonal billiards, appropriate modifications are necessary to take account of the fact that the flow is restricted to an invariant surface that is two dimensional. Further, since classical considerations do not always yield the spectrum $\{\lambda_n(t)\}$, we shall take resort to the semiclassical trace formula which involves periodic orbit sums similar to the kind that we shall encounter in the classical case.

Before considering the more general case of pseudo-integrable billiards, we first introduce the appropriate *classical* evolution operator for integrable systems. This is easily defined as

$$L^t \circ \phi(\theta_1, \theta_2) = \int \mathrm{d}\theta_1' \mathrm{d}\theta_2' \; \delta(\theta_1 - \theta_1'^t)\delta(\theta_2 - \theta_2'^t) \; \phi(\theta_1', \theta_2') \tag{4}$$

where θ_1 and θ_2 are the angular coordinates on the torus and evolve in time as $\theta_i^t = \omega_i(I_1, I_2)t + \theta_i$ with $\omega_i = \partial H(I_1, I_2)/\partial I_i$ and $I_i = (1/2\pi)\oint_{\Gamma_i} \mathbf{p}.\mathrm{d}\mathbf{q}$. Here $\Gamma_i, i = 1, 2$

refer to the two irreducible circuits on the torus and $\mathbf{p}$ is the momentum conjugate to the coordinate $\mathbf{q}$.

It is easy to see that the eigenfunctions $\{\phi_n(\theta_1,\theta_2)\}$ are such that $\phi_n(\theta_1^t,\theta_2^t) = \Lambda_n(t)\phi_n(\theta_1,\theta_2)$ where $\Lambda_n(t) = e^{i\alpha_n t}$. On demanding that $\phi_n(\theta_1,\theta_2)$ be a single valued function of (θ_1,θ_2), it follows that $\phi_{\mathbf{n}}(\theta_1,\theta_2) = e^{i(n_1\theta_1+n_2\theta_2)}$ where $\mathbf{n} = (n_1,n_2)$ is a point on the integer lattice. Thus the eigenvalue, $\Lambda_{\mathbf{n}}(t) = \exp\{it(n_1\omega_1 + n_2\omega_2)\}$.

To illustrate this, we consider a rectangular billiard where the hamiltonian expressed in terms of the actions, I_1, I_2 is $H(I_1,I_2) = \pi^2(I_1^2/L_1^2 + I_2^2/L_2^2)$ where L_1, L_2 are the lengths of the two sides. With $I_1 = \sqrt{E}L_1\cos(\varphi)/\pi$ and $I_2 = \sqrt{E}L_2\sin(\varphi)/\pi$, it is easy to see that at a given energy, E, each torus is parametrized by a particular value of φ. Thus

$$\Lambda_{\mathbf{n}}(t) = e^{i\,2\pi t\sqrt{E}(n_1\cos(\varphi)/L_1 + n_2\sin(\varphi)/L_2)} \tag{5}$$

and the spectrum is continuous. The trace thus involves an integration over φ as well as a sum over all n_1, n_2:

$$\mathrm{Tr}\, L^t = \sum_{\mathbf{n}} \int_{-\pi-\mu_n}^{\pi-\mu_n} \mathrm{d}\varphi\; e^{il\sqrt{E_{\mathbf{n}}}\sin(\varphi+\mu_n)} = 2\pi \sum_{\mathbf{n}} J_0(\sqrt{E_{\mathbf{n}}}\,l) \tag{6}$$

where J_0 is a Bessel function, $l = 2t\sqrt{E}$, $\tan(\mu_{\mathbf{n}}) = n_1L_2/(n_2L_1)$ and $E_n = \pi^2(n_1^2/L_1^2 + n_2^2/L_2^2)$. On separating out $\mathbf{n} = (0,0)$ from the rest and restricting the summation to the first quadrant of the integer lattice, it follows that

$$\mathrm{Tr}\, L^t = 2\pi + 2\pi N \sum_n J_0(\sqrt{E_n}\,l). \tag{7}$$

where $N = 4$. Note that the first term on the right merely states the fact that there exists a unit eigenvalue on every torus labelled by φ. Also, though we have not invoked semiclassics at any stage, the spectrum $\{E_n\}$ corresponds to the Neumann spectrum of the billiard considered. We shall subsequently show that this is true in general and for now it remains to express the trace of the evolution operator in terms of periodic orbits.

The trace of L^t in the integrable case can be expressed as

$$\mathrm{Tr}\, L^t = \int \mathrm{d}\varphi \int \mathrm{d}\theta_1\mathrm{d}\theta_2\, \delta(\theta_1 - \theta_1^t)\delta(\theta_2 - \theta_2^t). \tag{8}$$

It follows immediately that the only orbits that contribute are the ones that are periodic. For a rectangular billiard, the integrals can be evaluated quite easily and yields

$$\mathrm{Tr}\, L^t = 4\sum_{N_1}\sum_{N_2} \frac{4L_1L_2}{l_{N_1,N_2}}\,\delta(l - l_{N_1,N_2}) \tag{9}$$

where $\{N_1, N_2\}$ are the winding numbers on the torus and label periodic orbits of length l_{N_1,N_2}. Using (7) and (9), it follows that

$$4\sum_{N_1}\sum_{N_2} \frac{4L_1L_2}{l_{N_1,N_2}}\,\delta(l - l_{N_1,N_2}) = 2\pi + 2\pi N \sum_n J_0(\sqrt{E_n}\,l). \tag{10}$$

Thus, the dominant non-oscillatory contribution to the trace that survives averaging is 2π

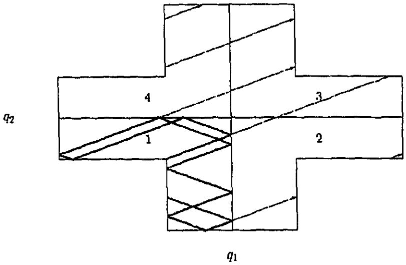


Figure 3. The singly connected region for an L-shaped billiard consists of four copies with edges appropriately identified. A trajectory originating near the $3\pi/2$ vertex in 1 is plotted in configuration space using bold lines and the corresponding unfolded trajectory is also shown. The latter consists of parallel segments and the trajectory can be parametrized by the angle φ that it makes for example with the q_1 axis.

and this gives rise to the analogue of (1)

$$4\left\langle \sum_{N_1} \sum_{N_2} \frac{4L_1L_2}{l_{N_1,N_2}}\,\delta(l - l_{N_1,N_2}) \right\rangle = 2\pi \tag{11}$$

The proliferation rate for the rectangular billiard thus follows from these considerations.

These ideas can be generalized for polygonal billiards that are pseudo-integrable even though the structure of the invariant surface no longer allows one to use *action* and *angle* variables.

For both integrable and pseudo-integrable polygonal billiards, the dynamics in phase space can be viewed in a singly connected region by executing $2g$ cuts in the invariant surface and identifying edges appropriately. At a given energy, the motion is parametrized by the angle, φ, that a trajectory makes with respect to one of the edges. As a trivial example, consider the rectangular billiard. The singly connected region is a larger rectangle consisting of four copies corresponding to the four directions that a trajectory can have and these can be glued appropriately to form a torus [12]. As a non-trivial example, consider the L-shaped billiard of figure (1) which is pseudo-integrable with its invariant surface having, $g = 2$. Alternately, the surface can be represented by a singly connected region in the plane (see figure 3) and consists of four copies corresponding to the four possible directions an orbit can have and these are glued appropriately. A trajectory in phase space thus consists of parallel segments at an angle φ measured for example with respect to one of the sides. It will be useful to note at this point that the same trajectory can also be represented by parallel segments at angles $\pi - \varphi$, $\pi + \varphi$ and $2\pi - \varphi$. In general, the number of directions for representing a trajectory equals the number of copies, N, that constitute the invariant surface.

The classical propagator on an invariant surface parametrized by φ is thus

$$L^t(\varphi) \circ \phi(\mathbf{q}) = \int \mathrm{d}\mathbf{q}'\,\delta(\mathbf{q} - \mathbf{q}'^t(\varphi))\,\phi(\mathbf{q}') \tag{12}$$

where $\mathbf{q}$ refers to the position in the singly connected region and $\mathbf{q}'^t(\varphi)$ is the time evolution parametrized by φ as described above.

The trace of $L^t(\varphi)$ takes into account all possible invariant surfaces that exist and hence involves an additional integration over φ. Thus

$$\mathrm{Tr}\, L^t = \int \mathrm{d}\varphi \int \mathrm{d}\mathbf{q}\, \delta(\mathbf{q} - \mathbf{q}^t(\varphi)). \tag{13}$$

Clearly the only orbits that contribute are the ones that are periodic. Further, the $\mathbf{q}$ integrations are simpler if we transform to a local coordinate system with one component parallel to the trajectory and the other perpendicular. Thus $\delta_{\|}(q_{\|} - q^t_{\|}) = (1/v)\delta(t - rT_p)$ where v is the velocity, T_p is the period of the orbit and r is the repetition number. Similarly, for an orbit of period T_p parametrised by the angle φ_p, $\delta_{\perp}(q_{\perp} - q^t_{\perp}) = \delta(\varphi - \varphi_p)/|\partial q_{\perp}/\partial\varphi|_{\varphi=\varphi_p}$ where $|\partial q_{\perp}/\partial\varphi|_{\varphi=\varphi_p} = rl_p$ for marginally unstable billiards. Putting these results together and noting that each periodic orbit occurs in general at N different values of φ, we finally have

$$\mathrm{Tr}\, L^t = \sum_n \Lambda_n(t) = N \sum_p \sum_{r=1}^{\infty} \frac{a_p}{rl_p}\, \delta(l - rl_p) \tag{14}$$

where $l = tv$ and the summation over p refers to all primitive periodic orbit families with length l_p and occupying an area a_p. Note that (9) is a special case of (14) which holds for both integrable and non-integrable polygonal billiards.

It is possible to re-express the periodic orbit sum in (14) starting with the appropriate quantum trace formula [13]

$$\sum_n \delta(E - E_n) = \mathrm{d}_{\mathrm{av}}(E) + \frac{1}{\sqrt{8\pi^3}} \sum_p \sum_{r=1}^{\infty} \frac{a_p}{\sqrt{krl_p}} \cos\left(krl_p - \frac{\pi}{4} - \pi r n_p\right) \tag{15}$$

Here $\mathrm{d}_{\mathrm{av}}(E)$ refers to the average density of quantal eigenstates, $k = \sqrt{E}$, l_p is the length of a primitive periodic orbit family. The phase $\pi r n_p$ is set to zero while considering the Neumann spectrum while in the Dirichlet case, n_p equals the number of bounces that the primitive orbit suffers at the boundary. For convenience, we have chosen $\hbar = 1$ and the mass $m = 1/2$. Starting with the function

$$g(l) = \sum_n f(\sqrt{E_n}\, l)\mathrm{e}^{-\beta E_n} = \int_{\epsilon}^{\infty} \mathrm{d}E\, f(\sqrt{E}\, l)\mathrm{e}^{-\beta E} \sum_n \delta(E - E_n) \tag{16}$$

where $f(x) = \sqrt{2/\pi x}\cos(x - \pi/4)$ and $0 < \epsilon < E_0$, it is possible to show using (15) that for polygonal billiards [15]

$$\sum_p \sum_{r=1}^{\infty} \frac{a_p}{rl_p}\, \delta(l - rl_p) = 2\pi b_0 + 2\pi \sum_n f(\sqrt{E_n}\, l) \tag{17}$$

for $\beta \to 0^+$. In the above,

$$b_0 = \sum_p \sum_r \frac{a_p(-1)^{rn_p}}{4\pi} \int_0^{\epsilon} \mathrm{d}E\, f(\sqrt{E}\, l) f(\sqrt{E}\, rl_p) \tag{18}$$

and is a constant [15, 16]. It follows from (17) and (14) that

$$\mathrm{Tr}\, L^t = \sum_p \sum_{r=1}^{\infty} \frac{Na_p}{rl_p}\, \delta(l - rl_p) = 2\pi N b_0 + 2\pi N \sum_n f(\sqrt{E_n}\, l). \tag{19}$$

billiard, the oscillatory contributions wash out on averaging so that

$$\left\langle \sum_{p} \sum_{r=1}^{\infty} \frac{a_p}{r l_p}\, \delta(l - r l_p) \right\rangle = 2\pi b_0 \tag{20}$$

This is the central result of this section and forms the basic sum rule obeyed by periodic orbit families.

A few remarks about this derivation and the magnitude of b_0 are however in order. Equation (14) is exact for the L-shaped billiard and all other boundaries which preclude the existence of isolated periodic orbits. However, even for these shapes, the semiclassical trace formula is only an approximation to the exact density whenever the billiard in question is pseudo-integrable. In this sense, the sum rule in (20) is not expected to be exact. However, we believe the existence of higher order corrections (such as diffraction) affects the magnitude of the constant b_0 while preserving the constant nature of the periodic orbit sum on the left.

In the integrable case it is easy to show from other considerations that $b_0 = 1/N$ where N is the number of sheets that constitute the invariant surface [15]. This also follows from (10) and (19) since $b_0 N = 1$. For pseudo-integrable billiards, each invariant surface parametrised by φ has an eigenvalue, $\Lambda_0(\varphi) = 1$. Thus the non-oscillatory part of the trace should equal $\int \Lambda_0(\varphi)\, \mathrm{d}\varphi$ and to a first approximation this yields 2π implying that $b_0 = 1/N$. However each singular vertex connects two distinct points at any angle φ and hence the integration over φ is non-trivial. We can therefore state that b_0 is approximately $1/N$ in the pseudo-integrable case while this is exact in the integrable case. We shall show numerically that the magnitude of deviations (from $1/N$) in the pseudo-integrable case depends on the existence of periodic orbit pairs at the singular vertex. First however, we briefly describe the algorithms used to determine periodic orbit families.

3. Algorithms for determining periodic orbits

Periodic orbits in polygonal billiards are generally hard to classify. Unlike integrable billiards, they cannot be described by two integers which count the number of windings around the two irreducible circuits on the torus though in exceptional cases this can indeed be done. However, since the invariant surface has a well-defined genus, it is expected that a set of integers $\mathrm{N} = \{N_1, N_2, \ldots, N_{2g}\}$ obeying the relationship

$$\omega_i = \frac{2\pi N_i}{T_{\mathrm{N}}} \tag{21}$$

can be used to label periodic orbits. Here ω_i refers to the frequency corresponding to each irreducible circuit Γ_i and depends on the energy, E and the angle φ that labels each invariant surface. Note however that not all points on this multi-dimensional integer lattice are allowed since there are constraints and this method of labelling orbits becomes cumbersome for surfaces of higher genus. Nevertheless, we illustrate the idea here for the L-shaped billiard of figure 1.

Let the length of the two bouncing ball orbits in the X-direction be L_1 and L_2 respectively and their lengths in the Y-direction be L_3 and L_4. These define the irreducible

$$\frac{}{2L_2} = \frac{}{T_{\mathrm{N}}}$$
$$\frac{v\sin(\varphi)}{2L_3} = \frac{2\pi N_3}{T_{\mathrm{N}}}$$
$$\frac{v\sin(\varphi)}{2L_4} = \frac{2\pi N_4}{T_{\mathrm{N}}} \qquad (22)$$

This implies that the angle φ at which a periodic orbit can exist is such that

$$\tan(\varphi) = \frac{N_3L_3 + N_4L_4}{N_1L_1 + N_2L_2}\,. \qquad (23)$$

Equations (23) and (22) merely express the fact that any periodic orbit should have integer number of windings around the irreducible circuits. Thus, the total displacement along the X-direction should be $2(N_1L_1 + N_2L_2)$ while the total displacement in the Y-direction should be $2(N_3L_3 + N_4L_4)$ where N_i are integers. As mentioned before however, not all realizations of $\{N_i\}$ correspond to real periodic orbits and the final step consists in checking numerically whether a periodic orbit at the angle φ (given by (23)) exists. Note that one member of each family necessarily resides at one of the singular vertices and hence it is sufficient to verify the existence of this orbit.

This method works equally well for other billiards with steps (see figure 1) and the number of integers necessary to describe orbits increases with the number of steps. An alternate method which exploits the fact that periodic orbits occur in families is often useful when the irreducible circuits are not obvious and this is described below.

Note that a non-periodic orbit originating from the same point **q** (e.g. the singular vertex) as a periodic orbit but with a momentum slightly different (from the periodic orbit) suffers a net transverse deviation $q_\perp = (-1)^{n_\varphi}\sin(\varphi - \varphi_p)l_\varphi$. Here l_φ is the distance traversed by a non-periodic orbit at an angle φ after n_φ reflections from the boundary and φ_p is the angle at which a periodic orbit exists. This provides a correction to the initial angle and a few iterations are normally sufficient to converge on a periodic orbit with good accuracy. In order to obtain all periodic orbits, it is necessary to shoot trajectories from every singular vertex since one member of each family resides at one of these vertices.

Apart from the length of a periodic orbit, it is also important to compute the area occupied by the family. This can be achieved by shooting a single periodic trajectory which resides at a singular vertex and by noting that this orbit lies on the edge of a family. Thus if the rest of the family lies on the left neighbourhood initially, it is necessary to determine the perpendicular distance from a singular vertex to this trajectory every time the initial neighbourhood lies towards this singular vertex. The shortest of these perpendicular distances gives the transverse extent of the family and the area can thus be computed.

These algorithms have been used to generate the lengths $\{l_p\}$ and areas $\{a_p\}$ of primitive orbits in the L-shaped billiard as well as the two and three step billiards. We present our numerical results in the following sections.

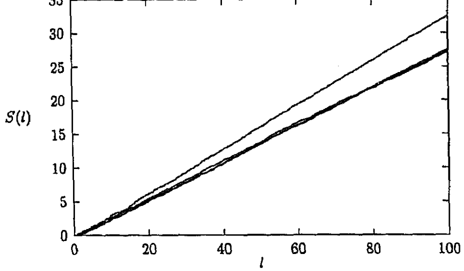


Figure 4. A demonstration of the sum rule for the 1-step billiard. The curve with slope significantly different corresponds to a degenerate example.

4. Numerical results: The sum rule and area law

We present here our numerical results on the 1-step (L-shaped), 2-step and 3-step billiards. Their invariant surfaces have genus 2, 3 and 4 respectively and the quantities we shall study are the sum rule derived in § 2 and the variation of $\langle a(l)\rangle$ with l and the genus, g of the invariant surface.

The sum rule we wish to study can be re-expressed as

$$S(l) = \frac{1}{2\pi}\sum_p \sum_{r} {}_{rl_p\le l} \frac{a_p}{rl_p} \simeq b_0 l \qquad (24)$$

and this is plotted in figure 4 for four different 1-step (L-shaped) billiards. Notice first that in each case the behaviour is linear as predicted by (24). Besides, in three of the four cases, the slopes are quite close ($b_0 \simeq 0.27$) and these correspond to non-degenerate L-shaped billiards with sides that are unequal and irrationally related. The one with a substantially larger slope ($b_0 \simeq 0.32$) is a degenerate case of a square with a quarter removed and for this system there exist a substantial number of periodic orbit pairs at the same angle on the two adjacent edges at the singular vertex. The differences between the degenerate and non-degenerate case also persists in other quantities as we shall shortly see.

We next plot $S(l)$ for non-degenerate examples of a 1, 2 and 3-step billiards in figure 5. The number of orbits considered are far less due to increased computational effort though the linear behaviour is obvious in all three cases. The slopes are again close to 0.25 and vary from $b_0 \simeq 0.24$ to $b_0 \simeq 0.28$.

Thus, periodic orbits obey a basic sum rule given by (24) where b_0 is a constant. Further, the magnitude of b_0 is close to 0.25 in all cases and the deviations from this value are larger in the degenerate case where periodic orbit pairs exist at the singular vertex.

The sum rule yields the proliferation rate of periodic orbits

$$N(l) = \frac{\pi b_0 l^2}{\langle a(l)\rangle} \qquad (25)$$

where $\langle a(l)\rangle$ is the average area occupied by periodic orbit families with length less than l. In the case of integrable polygons $\langle a(l)\rangle$ is a constant and equals AN where A is the area of the billiard while $b_0 = 1/N$. For pseudo-integrable cases, the areas a_p occupied by

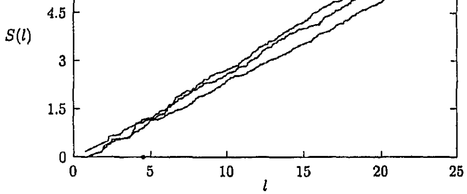


Figure 5. The sum rule for a 1, 2 and 3-step billiard. They are respectively the middle, bottom and top curves.

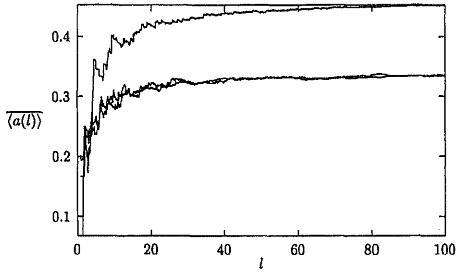


Figure 6. The normalized average area, $\overline{\langle a(l)\rangle} = \langle a(l)\rangle/NA$ as a function of the length l for 3 examples of the 1-step billiard. The saturation value differs for the degenerate example mentioned earlier.

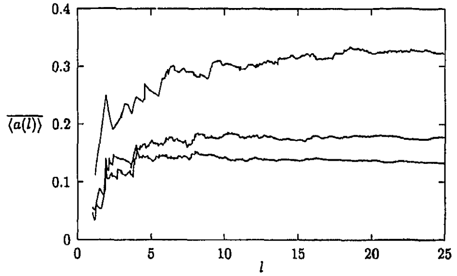


Figure 7. The normalized area for an example each of the 1-step (top), 2-step (middle) and 3-step (bottom) billiards.

individual periodic orbit families generically occupy a wide spectrum bounded above by NA. As the length of a family increases, encounters with singular vertices are more frequent so that the transverse extent of family decreases. Quantitative predictions about the behaviour of $\langle a(l)\rangle$ are however difficult and we provide here some numerical results.

Figure 6 shows a plot of $\langle a(l)\rangle/NA$ for three different L-shaped billiards one of which is the degenerate case presented in figure 4. The average area increases initially before saturating for large l in all cases. The normalized saturation value seems to be independent of the ratio of the sides as long as they are irrationally related (we have observed this for other cases not presented here) but is very different for the degenerate example of a square with a quarter removed.

A comparison of non-degenerate 1, 2 and 3-step billiards is shown in figure 7. Recall that the invariant surfaces for these have genus respectively equal to 2, 3 and 4. The saturation observed for the 1-step (L-shaped) case seems to be a common feature and interestingly the normalised average saturation value decreases. This is however expected since the number of singular vertices increases with the genus thereby reducing the average transverse extent of orbit families at any given length.

These observations allow us to conclude that for short lengths, the proliferation law is sub-quadratic. Asymptotically however, $N(l) \sim l^2$ as in integrable billiards. Further, the asymptotic proliferation rate for billiards with the same area A increases with the genus due to a decrease in the asymptotic normalized saturation value of $\langle a(l) \rangle$.

These numerical results provide a qualitative picture of the area law and show that periodic orbits in polygonal billiards are organized such that they obey a sum rule. Quantitative predictions would require a more extensive numerical study such that empirical laws for the saturation value and its variation with the genus can be arrived at.

5. Correlations in the length spectrum

Our numerical explorations so far have been focussed on the average properties of the length spectrum. We shall now attempt to understand the nature of the fluctuations and characterize their statistical properties.

Fluctuations in the length spectrum are generally difficult to study from purely classical considerations. There are notable exceptions however. Fluctuations in the integrable case can be studied using the Poisson summation formula since the lengths of orbits are expressed in terms of integers $\{N_1, N_2\}$ [17]. The other extreme is the motion on surfaces of constant negative curvature where the Selberg trace formula provides an *exact* dual relationship between the classical lengths and the eigenvalues of the Laplace-Beltrami operator [18]. For other systems, a possible way of studying fluctuations in the length spectrum lies in inverting the semiclassical quantum trace formula. For pseudo-integrable billiards, this has been achieved in § 2 and the integrable density of lengths, $N(l)$ can be expressed as

$$N(l) = \frac{\pi b_0 l^2}{\langle a(l) \rangle} + \frac{2\pi}{\langle a(l) \rangle} \sum_n \int_0^l \mathrm{d}l' \, l' f(\sqrt{E_n} l') \tag{26}$$

Statistical properties of the fluctuations can thus be studied using techniques introduced for the fluctuations in the quantum energy spectrum [19, 9].

The correlations commonly studied are the nearest neighbour spacings distribution, $P(s)$ and a two-point correlation referred to as the spectral rigidity, $\Delta_3(L)$. The rigidity measures the average mean square deviation of the staircase function $N(l)$ from the best fitting straight line over L mean level spacings. For a normalised (unit mean spacing) Poisson spectrum, $P(s) = \mathrm{e}^{-s}$ while $\Delta_3(L) = L/15$. These features are commonly observed in the quantum energy spectrum of integrable systems as well as in their length spectrum. For chaotic billiards, the quantum energy spectrum generically produces non-Poisson statistics while the length spectrum correlations are Poisson for long orbits at

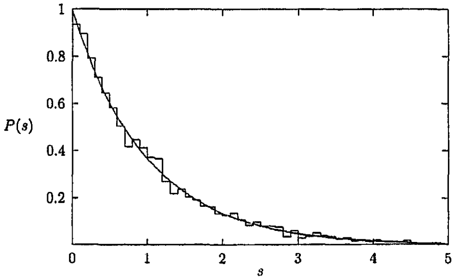


Figure 8. The nearest neighbour spacings distribution for a non-degenerate example of a 1-step billiard. Also plotted is the Poisson distribution.

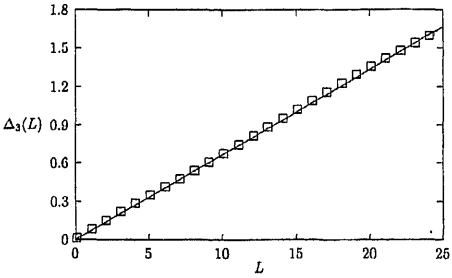


Figure 9. The spectral rigidity for the 1-step billiard together with the $L/15$ line.

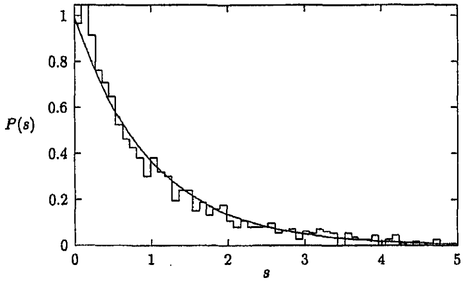


Figure 10. The nearest neighbour distribution for a 3-step billiard. The available lengths are much smaller and minor deviations from Poisson behaviour can be observed.

least over short ranges [20]. There are however deviations that can be observed in $\Delta_3(L)$ for short orbits or over longer ranges in the spectrum [20]. With this background, we now present our results for pseudo-integrable billiards.

Figures (8) and (9) show plots of $P(s)$ and $\Delta_3(L)$ for a non-degenerate L-shaped billiard. A total of 3000 lengths have been considered after excluding the shortest 3000 orbits. The correlations are clearly Poissonian as in the case of integrable or chaotic billiards.

For the 3-step billiard where fewer lengths are available (about 1250), we have carried out a similar study and the results are shown in figures (10) and (11). Deviations from the

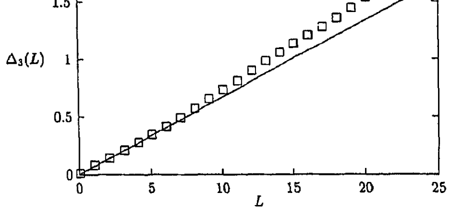


Figure 11. The spectral rigidity for the 3-step billiard considered earlier. The deviations from Poisson behaviour is more pronounced for this measure.

Possion behaviour can now be seen especially in the spectral rigidity. By considering shorter lengths in the L-shaped billiard, similar deviations were observed.

The statistical properties of fluctuations in the length spectrum of pseudo-integrable systems are thus similar to those of chaotic billiards. For long orbits the correlations are Poisson while deviations exist for shorter orbits.

6. Discussions and conclusions

Quantum billiards are experimentally realizable in the form of microwave cavities since the wave equations are identical [21, 22]. They are also interesting in their own right and have proved to be the testing grounds for several ideas in the field of quantum chaos. Of all possible shapes, polygonal billiards have perhaps been the least understood largely because very little was known about the organization of periodic orbits. The results presented here have however been used recently to obtain convergent semiclassical eigenvalues with arbitrary non-periodic trajectories [23] as well to demonstrate that periodic orbits provide excellent estimates of two-point correlations in the quantum energy spectrum [16]. Though it is beyond the scope of this article to review these recent developments on quantum polygonal billiards, we refer the interested reader to these articles and the references contained therein.

In the previous sections we have explored the organization of periodic orbits in rational polygonal billiards that are pseudo-integrable. Our main conclusions are as follows:

- Orbit families obey the sum rule $\langle \sum_p \sum_{r=1}^{\infty} a_p \delta(l - rl_p)/(rl_p) \rangle = 2\pi b_0$ thereby giving rise to the proliferation law $N(l) = \pi b_0 l^2 / \langle a(l) \rangle$ for all rational polygons.
- The quantity b_0 is approximately $1/N$ in generic rational billiards and deviations from this value are observed to be significant in degenerate situations.
- $\langle a(l) \rangle$ increases initially before saturating to a value much smaller than the maximum allowed area NA. The asymptotic proliferation law is thus quadratic even for systems that are not almost-integrable and the density of periodic orbits lengths is far greater than an equivalent integrable systems having the same area.
- The normalized average area $\langle a(l) \rangle / NA$ decreases with the genus of the billiard while b_0 is approximately $1/N$. Periodic orbits thus proliferate faster with an increase in genus.

- The statistical properties of the fluctuations in the length spectrum are Poisson when the orbits considered are long.

Acknowledgements

Part of the work reviewed here was published earlier in collaboration with Sudeshna Sinha and it a pleasure to acknowledge the numerous discussions we had on this subject. I have also benifitted directly or indirectly from interesting discussions with Predrag Cvitanović, Gregor Tanner and Bertrand Gerogeot.

References

[1] P J Richens and M V Berry, *Physica* **D2**, 495 (1981)
[2] The few examples of integrable polygons are the rectangle and triangles with internal angles $(\pi/3, \pi/3, \pi/3)$, $(\pi/3, \pi/6, \pi/2)$ and $(\pi/4, \pi/4, \pi/2)$
[3] M C Gutzwiller, *Chaos in classical and quantum mechanics* (Springer, New York, 1990)
[4] E Gutkin, *Physica* **D19**, 311 (1986); *J. Stat. Phys.* **83**, 7 (1996)
[5] H Masur, *Ergod. Th. Dyn. Sys.* **10**, 151 (1990); MSRI, 215 (1988)
[6] W A Veech, *Geo. Func. Anal.* **2**, 341 (1992)
[7] J Vega, T Uzer and J Ford, *Phys. Rev.* **E48**, 3414 (1993)
[8] T Cheon and T D Cohen, *Phys. Rev. Lett.* **62**, 2769 (1989)
[9] D Biswas and S Sinha, *Phys. Lett.* **A173**, 392 (1993)
[10] P Cvitanović and B Eckhardt, *J. Phys.* **A24**, L237 (1991)
[11] J H Hannay and A M Ozorio de Almeida, *J. Phys.* **A17**, 3429 (1984)
[12] J B Keller and S I Rubinow, *Ann. Phys. (NY)* **9**, 24 (1960)
[13] The respective contributions of isolated and diffractive periodic orbits are $O(k^{-1})$ and $O(k^{-1-\nu/2})$ where ν is the total number of (singular) vertex connections in a diffractive periodic orbit [14]
[14] N D Whelan, *Phys. Rev.* **E51**, 3778 (1995); N Pavloff and C Schmit, *Phys. Rev. Lett.* **75**, 61 (1995)
[15] D Biswas and S Sinha, *Phys. Rev. Lett.* **70**, 916 (1993)
[16] D Biswas, *Phys. Rev.* **E54**, R1044 (1996)
[17] D Biswas, *Pramana – J. Phys.* **42**, 447 (1994)
[18] A Selberg, *J. Indian Math. Soc.* **20**, 47 (1956)
N Balazs and A Voros, *Phys. Rep.* **143**, 109 (1986)
[19] M V Berry, *Proc. R. Soc. Lond. Ser.* **A400**, 229(1985)
[20] D Biswas, *Phys. Rev. Lett.* **71**, 2714 (1993)
[21] H-J Stockmann and J Stein, *Phys. Rev. Lett.* **64**, 2867 (1990)
[22] S Sridhar, *Phys. Rev. Lett.* **67**, 785 (1991)
[23] D Biswas, *Quasi-classical quantization*, to be published

Semiclassical theory for many-body fermionic systems

PIERRE GASPARD and SUDHIR R JAIN
Faculté des Sciences and Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Campus Plaine C.P. 231, Boulevard du Triomphe, B-1050 Brussels, Belgium

Abstract. We present a treatment of many-body fermionic systems that facilitates an expression of well-known quantities in a series expansion in $\hbar$. The ensuing semiclassical result contains, to a leading order of the response function, the classical time correlation function of the observable followed by the Weyl–Wigner series; on top of these terms are the periodic-orbit correction terms. The treatment given here starts from linear response assumption of the many-body theory and in its connection with semiclassical theory, it assumes that the one-body quantal system has a classically chaotic dynamics. Applications of the framework are also discussed.

Keywords. Fermionic systems; Weyl–Wigner series.

PACS Nos **05.30; 05.45**

1. Introduction

Semiclassical framework for single-particle systems is presently in an advanced stage [1, 2]. For both the integrable and chaotic dynamics, we now understand the semiclassical quantization. However, for intermediate [3] and mixed dynamical scenarios, we still lack a convincing theory. Another situation of a great practical relevance arises in the systems comprising of many bodies. That the spectral fluctuation characteristics of these systems are modelled in much the same way as for chaotic systems with lesser degrees of freedom has been shown quite recently [4]. It is very important to note that complex systems do possess features, mainly associated with the generic nature of the thermodynamic limit, which are completely absent from the systems with fewer freedoms, friction being one of the examples. There are other systems which one would like to understand semiclassically. Some notable examples are of quantum dots [5], metallic clusters [6] and their optical properties [7]; also, nuclear physics at high spins [8] presents us with opportunity to test various semiclassical ideas in many-body theory. There has been an attempt based on the semiclassical limit of the time-dependent Hartree–Fock equation where one can establish a connection [9] between the strength function given by the Vlasov equation and the corresponding quantum function in the limit of large quantum numbers. Analogous to the many-body treatment based on random-phase approximation (RPA), the linearized Vlasov equation gives an integral equation for the particle-hole propagator in terms of classical propagator in the static mean field and two-particle Coulomb interaction. Thus one needs to evaluate the classical propagator in the mean field and then solve the integral equation.

describe the shell effects in metallic clusters [10]. Of course, these have been preceded by nuclear physicists [11] by more than a decade. In fact, the periodic-orbit theory [12], which, in principle, should be rendered useless for practical purposes due to a huge number of technical problems associated with finding periodic orbits and summing a conditionally convergent series, gets encouragement from observations that only a few periodic orbits are enough in many situations [12].

We present here a semiclassical treatment of the response function. Response function is essentially the imaginary part of generalized dynamical susceptibility which, in turn, is related to the correlation commutator. In § 2, we present the general discussion from the linear response theory to arrive at a quantity for which semiclassical expression can be written. The key point of this section is to show that the most important quantity is a time correlation function. In § 3, we present semiclassical treatment of the system perturbed by an external influence. This enables us to write the relevant results of §§ 2 and 3 in a semiclassical expansion. We will see that the leading term is indeed the two-time correlation function averaged over the phase space, followed by the Weyl–Wigner series which has the periodic-orbit corrections. On our way, we take account of the fermionic nature of the particles.

2. Response function and the time correlation function

We present here a discussion on the response of a system in the presence of an external field. We work with susceptibility [13] and express it in terms of trace over single-particle states of two-time correlation function of the observable. This quantity is also known as the polarization propagator [14].

If a system described by the Hamiltonian $\hat{H}$ is externally disturbed by a field $F^{\mathrm{ext}}(t)$, then the total Hamiltonian is

$$\hat{H}_T = \hat{H} - \hat{Q}F^{\mathrm{ext}}(t), \tag{1}$$

where $\hat{Q}$ is an observable, an example could be magnetization in the context of spin systems, or, an electric dipole operator in an example involving photoabsorption, and so on. The response function can be written as the imaginary part of the dynamical susceptibility,

$$\begin{aligned} \chi''(t,t') &= (2\hbar)^{-1}\langle[\hat{Q}(t),\hat{Q}(t')]\rangle \\ &= \int \frac{d\omega}{2\pi}\, \mathrm{e}^{-i\omega(t-t')}\tilde{\chi}''(\omega). \end{aligned} \tag{2}$$

The angular brackets denote the expectation value and the square brackets denote the commutator. Setting an initial time to 0 and the final time to t, we can write

$$\chi''(t) = (2\hbar)^{-1}[\langle\hat{Q}(t)\hat{Q}(0)\rangle - \langle\hat{Q}(0)\hat{Q}(t)\rangle], \tag{3}$$

where

$$\hat{Q}(t) = \exp(i\hat{H}t/\hbar)\ \hat{Q}\ \exp(-i\hat{H}t/\hbar) \tag{4}$$

and where $\langle\cdot\rangle$ denotes the average over the initial state of the system which is, for instance, the thermal state wherein

$$\langle\cdot\rangle = \frac{1}{Z(\beta)}\,\mathrm{tr}\,\mathrm{e}^{-\beta\hat{H}}(\cdot) \tag{5}$$

with

$$Z(\beta) = \mathrm{tr}\exp(-\beta\hat{H}), \tag{6}$$

if the system was in contact with a thermal reservoir at temperature $T = 1/\beta k_B$ in the period preceding the interaction with the external field, $F^{\mathrm{ext}}(t)$. If the system was in its pure ground state $|\Phi_0\rangle$ then the average is over this many-body eigenstate of $\hat{H}$, i.e.,

$$\langle\cdot\rangle = \langle\Phi_0|\cdot|\Phi_0\rangle. \tag{7}$$

In this case, we can write $\chi''(t)$ as

$$\begin{aligned}\chi''(t) &= (2\hbar)^{-1}[\langle\Phi_0|\hat{Q}(t)\hat{Q}(0)|\Phi_0\rangle - \langle\Phi_0|\hat{Q}(0)\hat{Q}(t)|\Phi_0\rangle]\\ &= (2\hbar)^{-1}[\mathrm{e}^{iE_0t/\hbar}\langle\Phi_0|\hat{Q}\mathrm{e}^{-i\hat{H}t/\hbar}\hat{Q}|\Phi_0\rangle - \mathrm{e}^{-iE_0t/\hbar}\langle\Phi_0|\hat{Q}\mathrm{e}^{i\hat{H}t/\hbar}\hat{Q}|\Phi_0\rangle]\\ &= (2\hbar)^{-1}\sum_n[\mathrm{e}^{iE_0t/\hbar}|\langle\Phi_0|\hat{Q}|\Phi_n\rangle|^2\mathrm{e}^{-iE_nt/\hbar} - \mathrm{e}^{-iE_0t/\hbar}|\langle\Phi_0|\hat{Q}|\Phi_n\rangle|^2\mathrm{e}^{iE_nt/\hbar}],\end{aligned} \tag{8}$$

where $\{|\Phi_n\rangle\}$ denote all the many-body eigenstates of the isolated-system Hamiltonian,

$$\hat{H}|\Phi_n\rangle = E_n|\Phi_n\rangle, \tag{9}$$

with $E_n \geq E_0$.

Upon Fourier transformation,

$$\begin{aligned}\tilde{\chi}''(\omega) &= \pi\sum_n|\langle\Phi_0|\hat{Q}|\Phi_n\rangle|^2[\delta(\hbar\omega + E_0 - E_n) - \delta(\hbar\omega - E_0 + E_n)]\\ &= \sum_n|\langle\Phi_0|\hat{Q}|\Phi_n\rangle|^2\Im\left(\frac{1}{\hbar\omega - E_0 + E_n + i0^+} - \frac{1}{\hbar\omega + E_0 - E_n + i0^+}\right),\end{aligned} \tag{10}$$

where we used the identity

$$\frac{1}{x + i0^+} = \mathcal{P}\left(\frac{1}{x}\right) - i\,\pi\,\delta(x), \tag{11}$$

$\mathcal{P}$ denoting the Cauchy principal value. Methods have been developed to evaluate such expressions semiclassically [1]. However, we should recall that the system is many-body so that such methods would require the search for classical orbits of the many-body system. Simplification arises by taking account of the fermionic character of the system which allows a reduction of the problem to one-body Hamiltonian in an effective potential determined by, for instance, the time-independent Hartree–Fock method [6, 10, 11]. We shall make this simplifying assumption here and restrict our system to a set of uncoupled one-body Hamiltonians. A similar assumption is carried out on the coupling operator for which we assume the same form as for the

Hamiltonian,

$$\hat{H} = \sum_{i=1}^{N} \hat{h}_i,$$

$$\hat{Q} = \sum_{i=1}^{N} \hat{q}_i, \tag{12}$$

where $\hat{h}_i$ and $\hat{q}_i$ are one-body operators. Let us denote the one-body eigenstates as

$$\hat{h}|\phi_a\rangle = \epsilon_a|\phi_a\rangle. \tag{13}$$

When the fermionic many-body system is in its ground state, all the one-body eigenstates are occupied up to the Fermi energy, ϵ_F so that the energy of the ground state is

$$E_0 = \sum_{\substack{a=1 \\ \epsilon_a<\epsilon_F}}^{N} \epsilon_a. \tag{14}$$

In any many-body excited state, $|\Phi_n\rangle$, at least one one-body level above the Fermi energy is occupied. We may denote such excited states by the list of the occupied one-body levels, or equivalently, by the list of the one-body levels for which the occupation is different with respect to the ground state $|\Phi_0\rangle$, having in mind that such states are antisymmetric for an exchange of two fermions:

$$|\Phi_0\rangle = |1111...11\overset{\epsilon_F}{\vdots}\ 00...0\rangle$$

$$|\Phi_n\rangle = |111...1110\overset{\epsilon_F}{\vdots}\ 0100...\rangle. \tag{15}$$

For operators which are sums of one-body operators as assumed in (16), we obtain the result that the matrix elements $\langle\Phi_0|\hat{Q}|\Phi_n\rangle$ are non-vanishing only for the states Φ_n which differ from the ground state Φ_0 by one excitation. These states have a hole in state a ($\epsilon_a < \epsilon_F$) and a particle in state b ($\epsilon_b > \epsilon_F$). For these states the matrix elements are thus

$$\langle\Phi_0|\hat{Q}|\Phi_n\rangle = \langle\phi_a|\hat{q}|\phi_b\rangle. \tag{16}$$

Moreover the energy of the excited state is

$$E_n = E_0 - \epsilon_a + \epsilon_b \tag{17}$$

under the assumption that $\epsilon_a < \epsilon_F < \epsilon_b$. After some standard manipulations, we can write

$$\sum_n \frac{|\langle\Phi_0|\hat{Q}|\Phi_n\rangle|^2}{\hbar\omega - E_0 + E_n + i0^+}$$
$$= \sum_{a,b} \frac{|\langle\phi_a|\hat{q}|\phi_b\rangle|^2}{\hbar\omega - \epsilon_a + \epsilon_b + i0^+}\ [1 - \Theta(\epsilon_F - \epsilon_b)]\ \Theta(\epsilon_F - \epsilon_a), \tag{18}$$

where the Heaviside step function, $\Theta(\cdot)$ takes care of the aforementioned restriction on the location of the one-body states ϕ_a and ϕ_b with respect to ϵ_F.

In the presence of a thermal reservoir at temperature $T[= 1/(\beta k_B)]$ described with canonical density matrix, we get that

$$\begin{aligned}\langle \hat{Q}(t)\hat{Q}(0)\rangle &= \frac{1}{Z}\ \mathrm{tr} e^{-\beta\hat{H}} e^{it\hat{H}/\hbar}\hat{Q}\ e^{-it\hat{H}/\hbar}\hat{Q} \\ &= \sum_{mn} \exp[iE_m(t+i\beta\hbar)/\hbar]\ \exp(-iE_n t/\hbar)\ |\langle\Phi_m|\hat{Q}|\Phi_n\rangle|^2. \end{aligned} \tag{19}$$

Thus the dynamical susceptibility is

$$\tilde{\chi}''(\omega) = \frac{1}{Z(\beta)}\sum_{mn} |\langle\Phi_m|\hat{Q}|\Phi_n\rangle|^2\ \Im\frac{\exp(-\beta E_m)}{\hbar\omega - E_m + E_n + i0^+} - (\omega \to -\omega). \tag{20}$$

Similar assumptions as before enable us to reduce this expression to the one-body system. Now the probability to find a state $|\Phi_m\rangle$ in which $|\phi_a\rangle$ is occupied is given by the Fermi-Dirac distribution at energy ϵ_a. But $|\Phi_n\rangle$ is related to $|\Phi_m\rangle$ by the fact that $|\phi_b\rangle$ must be unoccupied (see above). So, we have a joint probability that $|\phi_a\rangle$ is occupied and $|\phi_b\rangle$ is unoccupied, hence the propagator is

$$\tilde{\chi}''(\omega) = \sum_{a,b} |\langle\phi_a|\hat{q}|\phi_b\rangle|^2\ \Im\frac{p_a^{FD} - p_b^{FD}}{\hbar\omega - \epsilon_a + \epsilon_b + i0^+}, \tag{21}$$

where

$$p_a^{FD} = \frac{1}{\exp[\beta(\epsilon_a - \mu)] + 1}, \tag{22}$$

denotes the Fermi–Dirac probability or mean occupation number. μ is the chemical potential fixed by the total number of fermions in the system. Equation (21) gives therefore the Fourier transform of the two-time correlation function.

At zero temperature, the Fermi–Dirac distribution becomes a Heaviside step function, so the expression (21) becomes

$$\tilde{\chi}''(\omega) = \sum_{a,b} |\langle\phi_a|\hat{q}|\phi_b\rangle|^2\ \Im\frac{\Theta(\epsilon_F - \epsilon_a) - \Theta(\epsilon_F - \epsilon_b)}{\hbar\omega - \epsilon_a + \epsilon_b + i0^+}, \tag{23}$$

which is identical with the expression previously derived. In the foregoing discussion of this section, we have started from the time correlation function (actually the commutator) and written an expression for the response. Now that the susceptibility has been reduced to a 1-body expression, we can treat the 1-body system semiclassically in terms of periodic orbits in the effective Hartree–Fock Hamiltonian $\hat{h}$.

It is well to recall that $\tilde{\chi}''(\omega)$ has a physical interpretation in terms of energy dissipation which is related to the fluctuations in the frequency spectrum obtained as the Fourier transform of the time correlation function. Indeed the fluctuation-dissipation theorem [13] lies just in realizing that connection.

Employing the identity (11) in (21), and expanding the difference between p^{FD}'s in the resulting expression, we obtain

$$\begin{aligned}\tilde{\chi}''(\omega) &= \pi \sum_{a,b} |\langle\phi_a|\hat{q}|\phi_b\rangle|^2 \, [p^{FD}(\epsilon_a) - p^{FD}(\epsilon_a + \hbar\omega)] \, \delta(\hbar\omega + \epsilon_a - \epsilon_b) \\ &= -\pi \sum_{n=1}^{\infty} \frac{(\hbar\omega)^n}{n!} \sum_{a,b} \frac{\partial^n p^{FD}}{\partial\epsilon^n}(\epsilon_a) |\langle\phi_a|\hat{q}|\phi_b\rangle|^2 \, \delta(\hbar\omega + \epsilon_a - \epsilon_b) \\ &= -\frac{1}{2\hbar} \sum_{n=1}^{\infty} \frac{(\hbar\omega)^n}{n!} \tilde{f}_n(\omega), \end{aligned} \tag{24}$$

where we introduced the expression

$$\begin{aligned}\tilde{f}_n(\omega) &= 2\pi\hbar \sum_{a,b} \frac{\partial^n p^{FD}}{\partial\epsilon^n}(\epsilon_a) \, |\langle\phi_a|\hat{q}|\phi_b\rangle|^2 \delta(\hbar\omega + \epsilon_a - \epsilon_b) \\ &= \int \mathrm{d}t \, \mathrm{e}^{i\omega t} f_n(t), \end{aligned} \tag{25}$$

which is the Fourier transform of the time correlation functions

$$f_n(t) = \mathrm{tr} \frac{\partial^n p^{FD}}{\partial\epsilon^n}(\hat{h}) \mathrm{e}^{it\hat{h}/\hbar} \hat{q} \mathrm{e}^{-it\hat{h}/\hbar} \hat{q}. \tag{26}$$

Moreover, we can rewrite these time correlation functions in terms of a single time correlation function as follows

$$f_n(t) = \int \mathrm{d}\epsilon \frac{\partial^n p^{FD}}{\partial\epsilon^n}(\epsilon) \; C_\epsilon(t), \tag{27}$$

with

$$C_\epsilon(t) = \mathrm{tr} \, \delta(\epsilon - \hat{h}) \; \hat{q}(t) \; \hat{q}(0) \qquad \text{and} \qquad \hat{q}(t) = \mathrm{e}^{it\hat{h}/\hbar} \hat{q} \mathrm{e}^{-it\hat{h}/\hbar}. \tag{28}$$

In summary, we have so far reduced the many-body dynamical susceptibility to an expression involving the time correlation function $C_\epsilon(t)$ of the 1-body effective dynamics:

$$\tilde{\chi}''(\omega) = -\pi \sum_{n=1}^{\infty} \frac{(\hbar\omega)^n}{n!} \int \frac{\mathrm{d}t\mathrm{d}\epsilon}{2\pi\hbar} \; \mathrm{e}^{i\omega t} \; \frac{\partial^n p^{FD}}{\partial\epsilon^n}(\epsilon) \; C_\epsilon(t). \tag{29}$$

We will see in the next section that this expression is particularly suited for developing semiclassical expansions.

3. Semiclassical expressions of response

In many cases, fermions occupy the one-body levels up to high quantum numbers. Accordingly, a semiclassical method (like for instance WKB) turns out to be justified, in which the quantum-mechanical quantities are obtained as asymptotic expansions in terms of classical mechanics. However, progress in the last decades has shown that chaotic motion is a typical feature of classical mechanics so that an exponential proliferation of periodic orbits is to be expected. The classical dynamics in the three-dimensional one-body effective potential of the fermionic system will in general present chaotic motion because chaos is possible in systems with more than two degrees of freedom. Chaotic motion is often mixed with quasiperiodic motion, specially, at low energies. At higher energies, chaotic motion becomes dominant in many systems, which justifies the assumption used in the following that all the periodic orbits are unstable and isolated like in fully developed chaos. It is under this assumption of a chaotic dynamical scenario that Gutzwiller derived his famous semiclassical trace formula [12]. Our purpose is now to apply Gutzwiller's method in order to obtain the response function semiclassically.

To this end, we need the semiclassical expression for the time correlation function $C_\epsilon(t)$, which we consider in the form:

$$C_\epsilon(t) = \operatorname{tr}\delta(\epsilon-\hat{h})\hat{X} = \operatorname{tr}\delta(\epsilon-\hat{h})\hat{A} + i\operatorname{tr}\delta(\epsilon-\hat{h})\hat{B}, \tag{30}$$

where $\hat{X} = \hat{X}(t) = \hat{q}(t)\hat{q}(0)$ is a one-body operator. This operator is non-hermitian but can be decomposed as $\hat{X} = \hat{A} + i\hat{B}$ in terms of two Hermitian operators: $\hat{A} = (\hat{X}+\hat{X}^\dagger)/2$ and $\hat{B} = (\hat{X}-\hat{X}^\dagger)/(2i)$.

Methods have been obtained to evaluate such expressions semiclassically in terms of periodic orbits [1, 15]. To motivate this method, we begin by observing [16] that such an expression involves the matrix elements of $\hat{A}$ (and $\hat{B}$) over the eigenstates of $\hat{h}$. These matrix elements can be obtained at the first-order perturbation theory of a perturbed Hamiltonian $\hat{h}(\lambda) = \hat{h} + \lambda\hat{A}$. Assuming the eigenvalue problem for $\hat{h}(\lambda)$ to be solved, the matrix elements of $\hat{A}$ may thus be obtained in terms of derivatives of eigenvalues of the perturbed Hamiltonian, $\epsilon_n(\lambda)$, with respect to λ. The periodic-orbit theory by Gutzwiller can then be used to calculate the diagonal matrix elements.

Each term of the correlation function $C_\epsilon(t)$ can therefore be expressed as

$$\begin{aligned}\operatorname{tr}\delta(\epsilon-\hat{h})\hat{A} &= -\frac{1}{\pi}\Im\operatorname{tr}\frac{\hat{A}}{\epsilon-\hat{h}+i0^+}\\ &= \frac{1}{\pi}\Im\operatorname{tr}\frac{\partial}{\partial\lambda}\log(\epsilon-\hat{h}-\lambda\hat{A}+i0^+)\Big|_{\lambda=0}.\end{aligned} \tag{31}$$

Comparing with the following identity

$$\begin{aligned}-\frac{1}{\pi}\frac{\partial}{\partial\epsilon}\Im\operatorname{tr}\log(\epsilon-\hat{h}-\lambda\hat{A}+i0^+) &= \operatorname{tr}\delta(\epsilon-\hat{h}-\lambda\hat{A})\\ &= \frac{\partial}{\partial\epsilon}N(\epsilon;\lambda),\end{aligned} \tag{32}$$

which is the derivative of the staircase function $N(\epsilon;\lambda)$ with respect to the energy, we arrive at the relation,

$$\operatorname{tr}\delta(\epsilon-\hat{h})\hat{A} = -\frac{\partial N(\epsilon;\lambda)}{\partial\lambda}\Big|_{\lambda=0}, \tag{33}$$

which gives each term of the correlation function as the derivative of the staircase periodic orbits are unstable and isolated:

$$N(\epsilon;\lambda)=\int\frac{\mathrm{d}^f x\mathrm{d}^f p}{(2\pi\hbar)^f}\,\Theta[\epsilon-h_W(\lambda)]+O(\hbar^{-f+1})$$
$$+\sum_p\sum_{r=1}^{\infty}\frac{1}{r\pi}\frac{\sin[\frac{r}{\hbar}S_p(\epsilon;\lambda)-r\frac{\pi}{2}\nu_p]}{|\det[m_p^r(\lambda)-I]|^{1/2}}+O(\hbar), \tag{34}$$

where $m_p(\lambda)$ is the monodromy matrix governing the stability of the classical periodic trajectory, p; ν_p is the Maslov index of the trajectory and h_W is the Weyl–Wigner transform of the Hamiltonian $\hat{h}(\lambda)=\hat{h}+\lambda\hat{A}$. The Gutzwiller-type expression (34) contains different expansions. First of all, we find series in powers of $\hbar$ (such as the Weyl corrections to the average staircase function). These series, which also appear in the WKB method, are known to be asymptotic series but with a controllable rest in general. The properties of these series in $\hbar$ have been discussed in the literature [19–22]. On the other hand, several rigorous results exist about the convergence of the sum over periodic orbits [19, 20, 22]. Different regularizations have been studied which are based on Lorentzian smoothing (i.e., on the complexification of energy [23]) or on Gaussian smoothing [24, 25] and which yield results with controllable error bounds.

We notice that the above periodic orbits p in equation (34) are those of the perturbed Hamiltonian. We assume here that the periodic orbits of $h_W(\lambda)$ deform continuously to the periodic orbits of $h_W(\lambda=0)$ for λ small enough. For the purpose of differentiating this expression with respect to λ, we use the following classical formula which gives the derivative of the action of the periodic orbits [26],

$$\frac{\partial S_p}{\partial\lambda}=-\oint\mathrm{d}t\,\frac{\partial h_W(\lambda)}{\partial\lambda}=-\oint\mathrm{d}t\,A_W, \tag{35}$$

where the integrals go around the periodic orbit, p, and A_W is the Weyl–Wigner transform of the operator $\hat{A}$. Taking the derivative with respect to λ and adding both terms composing the correlation function, we finally obtain

$$\mathrm{tr}\,\delta(\epsilon-\hat{h})\hat{X}=\int\frac{\mathrm{d}^f x\mathrm{d}^f p}{(2\pi\hbar)^f}\,X_W\,\delta(\epsilon-h_{cl})+O(\hbar^{-f+1})$$
$$+\frac{1}{\pi\hbar}\sum_p\sum_{r=1}^{\infty}\frac{\cos(\frac{r}{\hbar}S_p-r\frac{\pi}{2}\nu_p)}{|\det(m_p^r-I)|^{1/2}}\oint_p\mathrm{d}t\,X_W+O(\hbar^0) \tag{36}$$

where $h_{cl}=h_W(\lambda=0)$ and the periodic orbits are those of h_{cl}.

For the case like ours, when

$$\hat{X}=\exp\left(\frac{i}{\hbar}\hat{h}t\right)\hat{q}\exp\left(-\frac{i}{\hbar}\hat{h}t\right)\hat{q}, \tag{37}$$

the Weyl–Wigner tranform can be written as

$$X_W(\mathbf{x},\mathbf{p})=(\mathrm{e}^{\frac{i}{\hbar}\hat{h}t})_W\mathrm{e}^{\frac{i}{2}\hbar\hat{\Lambda}}\{q_W\,\mathrm{e}^{\frac{i}{2}\hbar\hat{\Lambda}}[(\mathrm{e}^{\frac{-i}{\hbar}\hat{h}t})_W\mathrm{e}^{\frac{i}{2}\hbar\hat{\Lambda}}q_W]\}$$

$$= [\exp(-\hat{\mathcal{L}}_{cl}t)q_W]q_W + (\text{Weyl corrections})$$

$$= q_W(t)q_W(0) + (\text{Weyl corrections}), \tag{38}$$

ere $\hat{\mathcal{L}}_{cl} = \{h, \cdot\}$ is the classical evolution operator, and the operator $\hat{\Lambda}$ is standard [27] the Weyl–Wigner expansions defined as

$$\hat{\Lambda} = \frac{\overleftarrow{\partial}}{\partial \mathbf{p}} \cdot \frac{\overrightarrow{\partial}}{\partial \mathbf{x}} - \frac{\overleftarrow{\partial}}{\partial \mathbf{x}} \cdot \frac{\overrightarrow{\partial}}{\partial \mathbf{p}}, \tag{39}$$

d $q_W(t) = q_W[\Phi^t(\mathbf{x}, \mathbf{p})]$ is the value of q_W at the current point $\Phi^t(\mathbf{x}, \mathbf{p})$ of the trajectory m the initial condition $(\mathbf{x}, \mathbf{p})$ of the Hamiltonian flow Φ^t of h_{cl}.

Putting all the pieces together, we can now write the semiclassical expansion for the rrelation function as

$$\begin{aligned}
C_\epsilon(t) &= \operatorname{tr} \delta(\epsilon - \hat{h})\hat{q}(t)\hat{q}(0) \\
&= \int \frac{\mathrm{d}^f x \mathrm{d}^f p}{(2\pi\hbar)^f}\ \delta[\epsilon - h_{cl}(\mathbf{x}, \mathbf{p})]\ q_W(\mathbf{x}, \mathbf{p})(\mathrm{e}^{-\hat{\mathcal{L}}_{cl}t} q_W)(\mathbf{x}, \mathbf{p}) \\
&\quad + (\text{Weyl corrections}) \\
&\quad + \frac{1}{\pi\hbar} \sum_{p,r} \frac{\cos(\frac{r}{\hbar}S_p - r\frac{\pi}{2}\nu_p)}{|\det(m_p^r - I)|^{1/2}} \oint_p \mathrm{d}\tau\, q_W(\tau) q_W(\tau + t) + O(\hbar^0).
\end{aligned} \tag{40}$$

e now go back to equation (29) where we see that we still need to evaluate the Fourier nsform of the correlation function $C_\epsilon(t)$.

Let us evaluate the first term with $n = 1$ of the susceptibility. This should give the minant term of the Weyl series in powers of the Planck constant $\hbar$ as well as a sum over riodic orbits with the smallest power in $\hbar$ which gives therefore the most important ntribution. The other periodic-orbit sums would involve smaller amplitudes with higher wers in $\hbar$. Replacing with the previous semiclassical result, we get

$$\begin{aligned}
\tilde{\chi}''(\omega) &= -\frac{\omega}{2} \int \mathrm{d}\epsilon\ \frac{\partial p^{FD}}{\partial \epsilon}(\epsilon) \int \mathrm{d}t\ \mathrm{e}^{i\omega t}\ C_\epsilon(t) + \mathcal{O}(\hbar) \\
&= -\frac{\omega}{2} \int \mathrm{d}t \int \frac{\mathrm{d}^f x \mathrm{d}^f p}{(2\pi\hbar)^f}\ \partial_\epsilon p^{FD}[h_{cl}(\mathbf{x}, \mathbf{p})] q_W(\mathbf{x}, \mathbf{p}) \\
&\quad \times [\mathrm{e}^{(i\omega - \hat{\mathcal{L}}_{cl})t} q_W](\mathbf{x}, \mathbf{p}) + \mathcal{O}(\hbar^{-f+1}) \\
&\quad - \frac{\omega}{2\pi\hbar} \int \mathrm{d}\epsilon \partial_\epsilon p^{FD}(\epsilon) \sum_{p,r} \frac{\cos(\frac{r}{\hbar}S_p - r\frac{\pi}{2}\nu_p)}{|\det(m_p^r - I)|^{1/2}} \int \mathrm{d}t\, \mathrm{e}^{i\omega t} \\
&\quad \times \oint_p \mathrm{d}\tau q_W(\tau) q_W(\tau + t) + O(\hbar^0).
\end{aligned} \tag{41}$$

ssuming that $\hat{q} = \hat{q}^\dagger$ is a Hermitian operator, its Weyl-Wigner transform is a real nction of $(\mathbf{x}, \mathbf{p})$. Along a periodic orbit, we have the Fourier series

$$q_W(\tau) = \sum_{n=-\infty}^{+\infty} q_{p,n} \exp\left(i \frac{2\pi n}{T_p} \tau\right), \tag{42}$$

where T_p is the classical period and $q_{p,-n} = q^*_{p,n}$. We obtain that

$$\int \mathrm{d}t\, \mathrm{e}^{i\omega t} \oint_p \mathrm{d}\tau q_W(\tau)\, q_W(\tau + t) = 2\pi\, T_p(\epsilon) \sum_{n=-\infty}^{+\infty} |q_{p,n}(\epsilon)|^2\, \delta\left[\omega - \frac{2\pi n}{T_p(\epsilon)}\right]. \tag{43}$$

Except in special systems like the harmonic oscillators, the periods $T_p(\epsilon)$ vary with the energy ϵ. Because of the presence of the Dirac distribution, we can perform the last integral over ϵ by using

$$\delta\left[\omega - \frac{2\pi n}{T_p(\epsilon)}\right] = \frac{T_p^2}{2\pi n \partial_\epsilon T_p}\, \delta[\epsilon - \epsilon_{p,n}(\omega)], \tag{44}$$

where $\epsilon = \epsilon_{p,n}(\omega)$ is the energy obtained by solving the implicit equation $T_p(\epsilon) = 2\pi n/\omega$ (when this solution exists). We here assumed that the periods vary monotonically with energy, otherwise we find a sum of terms instead of a single term in the right-hand member of the previous formula. Noting that

$$\frac{\mathrm{d}S_{p,n}}{\mathrm{d}\omega} = \frac{\mathrm{d}}{\mathrm{d}\omega} S_p[\epsilon_{p,n}(\omega)] = -\frac{T_p^3}{2\pi n \partial_\epsilon T_p}, \tag{45}$$

we have that

$$\int \mathrm{d}t\, \mathrm{e}^{i\omega t} \oint_p \mathrm{d}\tau\, q_W(\tau) q_W(\tau + t) = 2\pi \sum_n |q_{p,n}(\epsilon)|^2 \left|\frac{\mathrm{d}S_{p,n}}{\mathrm{d}\omega}\right| \delta[\epsilon - \epsilon_{p,n}(\omega)], \tag{46}$$

where $S_{p,n} = S_p[\epsilon_{p,n}(\omega)]$.

Substituting in the previous relation, we finally obtain the absorptive part of the dynamical susceptibility as:

$$\begin{aligned}\tilde{\chi}''(\omega) = &-\frac{\omega}{2} \int \mathrm{d}t \int \frac{\mathrm{d}^f x \mathrm{d}^f p}{(2\pi\hbar)^f}\, \partial_\epsilon p^{FD}[h_{cl}(\mathbf{x},\mathbf{p})] q_W(\mathbf{x},\mathbf{p}) \\ &\times [\mathrm{e}^{(i\omega - \hat{\mathcal{L}}_{cl})t} q_W](\mathbf{x},\mathbf{p}) + \mathcal{O}(\hbar^{-f+1}) \\ &- \frac{\omega}{\hbar} \sum_{p,r,n} \partial_\epsilon p^{FD}\big|_{p,n}\, |q_{p,n}|^2 \left|\frac{\mathrm{d}S_{p,n}}{\mathrm{d}\omega}\right| \frac{\cos(\frac{r}{\hbar} S_{p,n} - r\frac{\pi}{2}\nu_p)}{|\det(m^r_{p,n} - I)|^{1/2}} + \mathcal{O}(\hbar^0),\end{aligned} \tag{47}$$

where all the expressions with indices p, n are functions of the frequency ω through their dependency on the energy $\epsilon = \epsilon_{p,n}(\omega)$. Equation (47) is the central result of this work.

The formula (47) has delicate convergence properties which are similar to those of Gutzwiller trace formula, as discussed above. Such periodic-orbit and $\hbar$ expansions have already been applied to several systems where reliable and also high-precision calculations can be carried out with appropriate truncations [1, 2, 15, 17, 26, 28]. To interpret the result (47), let us go back to the time correlation function $C_\epsilon(t)$ which is

$$\begin{aligned} &\operatorname{tr} \delta(\epsilon - \hat{h}) \exp(i\hat{h}t/\hbar)\hat{q} \exp(-i\hat{h}t/\hbar)\hat{q} \\ &= \frac{1}{2\pi} \int \mathrm{d}\tau \, \mathrm{e}^{i\epsilon\tau/\hbar} \operatorname{tr} \mathrm{e}^{i\hat{h}(t-\tau)/\hbar} \, \hat{q} \, \mathrm{e}^{-i\hat{h}t/\hbar} \, \hat{q}. \end{aligned} \tag{48}$$

In this expression, the trace can be rewritten in terms of the Green propagators as

$$\begin{aligned} &\operatorname{tr} \exp[i\hat{h}(t-\tau)/\hbar]\hat{q} \exp[-i\hat{h}t/\hbar]\hat{q} \\ &= \int \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{x}' G(\mathbf{x}', \mathbf{x}; \tau - t) q(\mathbf{x}) G(\mathbf{x}, \mathbf{x}'; t) q(\mathbf{x}') \\ &= \sum_{c,c'} \int \mathrm{d}\mathbf{x} \, \mathrm{d}\mathbf{x}' A_c(\mathbf{x}', \mathbf{x}; \tau - t) \exp[iW_c(\mathbf{x}', \mathbf{x}; \tau - t)/\hbar] q(\mathbf{x}) \\ &\quad \times A_{c'}(\mathbf{x}, \mathbf{x}'; t) \exp[iW_{c'}(\mathbf{x}, \mathbf{x}'; t)/\hbar] q(\mathbf{x}'), \end{aligned} \tag{49}$$

where in the last step, we have written down symbolically the semiclassical expressions for the Green propagators. We have here assumed that the operator $\hat{q}$ depends only on the position operator, i.e., $\hat{q} = q(\hat{\mathbf{x}})$. The sum in the above expressions is over classical paths, c and c'. Notice that these are two different paths in general as they arise from two different propagators. The expression interprets for itself — from the point $\mathbf{x}$ to $\mathbf{x}'$, we propagate along the path c' for time t, and then we propagate back to the point $\mathbf{x}$ along c in time $\tau - t$. Precisely,

$$\begin{aligned} W_c(\mathbf{x}' \to \mathbf{x}; \tau - t) &= \int_0^{\tau - t} \mathrm{d}t \, L \\ W_{c'}(\mathbf{x} \to \mathbf{x}'; t) &= \int_0^{t} \mathrm{d}t \, L, \end{aligned} \tag{50}$$

L being the Lagrangian of the system. The total phase in the correlation function is the sum $W_c + W_{c'}$ of the two actions. By stationary phase method, it can be shown that the two segments of the paths c and c', in fact, form a periodic orbit of period τ. This is the mechanism by which periodic orbit corrections appear in the main result above.

Classical contribution comes from the paths of zero length, i.e., $\tau = 0$, and this is precisely the leading term in our expression, also called the quasiclassical term. Thus, in any calculation of interest, we need to find the classical correlation function, the Weyl corrections, and then the periodic-orbit corrections. If the energy spectrum of the system is discrete, we notice that the dynamical susceptibility (24) is a sum of Dirac peaks centered on each of the Bohr frequencies $\omega = (E_m - E_n)/\hbar$. In the semiclassical formula (47) the first term due to the paths of zero length contribute to the dynamical susceptibility by a continuous function of frequency ω which may be interpreted as the average or background of the discrete quantum susceptibility. The second series of terms of (47) are due to the periodic orbits which contribute by oscillatory functions of the frequency ω. Their sum is supposed to reproduce the Dirac peaks of the discrete quantum susceptibility $\tilde{\chi}''(\omega)$.

We notice that in contrast with the well-known Gutzwiller formula for the level density, the quasiclassical term here involves a classical time correlation function. As a consequence, the spectral properties of the classical Liouvillian $\hat{\mathcal{L}}_{cl}$ will intervene in the evaluation of this term (as well as of the following Weyl corrections). In this regard, we mention that a periodic-orbit theory of the classical Liouvillian has been developed [17, 18] in terms of classical zeta functions which resemble (but differ from) the Selberg-Gutzwiller quantum zeta function. Consequently, we have here an example where the semiclassical evaluation of a quantum property, namely $\tilde{\chi}''(\omega)$, involves the spectral properties of the classical Liouvillian along with the Gutzwiller quantum periodic-orbit sum. It is remarkable that both these classical and semiclassical terms coexist within the same formula, which points to the result that the classical periodic-orbit theory is only one part of the underlying quantum dynamics. We believe that this result is general in the semiclassical evaluation of dynamical susceptibilities, which include, in particular, the ac conductivity.

It turns out that the spectral properties of the Liouvillian differ considerably whether the classical system h_{cl} is integrable or chaotic.

For integrable systems, the Liouvillian spectrum is formed by discrete classical frequencies, $\hat{\mathcal{L}}_{cl}\psi_{\mathbf{m}} = i\Omega_{\mathbf{m}}\psi_{\mathbf{m}}$, on each ergodic component which are the invariant tori. The expression would then be similar to the one obtained in [7]. The periodic-orbit corrections should be given by the Berry–Tabor semiclassical theory under such circumstances [29]. Let us mention that a trace formula has been recently derived for the case where an integrable system is deformed so that a continuous symmetry becomes approximate [30].

For chaotic systems, the classical time correlation function in the quasiclassical term decays if the system is mixing on the energy shell. Recent works have shown that this decay may be developed in terms of the complex singularities of an analogue of the Liouvillian resolvent [17, 18, 31]. When the complex singularities are poles they are called Pollicott–Ruelle resonances and are associated with exponential decays. If the spectral decomposition of the Liouvillian dynamics in terms of the Pollicott–Ruelle resonances is known the quasiclassical term can be further reduced as a sum over the Pollicott–Ruelle resonances over each energy shell [17, 18, 31].

4. Concluding remarks

We have described a systematic procedure which lets us write important quantities in many-body theory in a semiclassical expansion. All our considerations are restricted to many-body systems reducible to uncoupled one-body systems. We believe that it is an interesting first step. We have shown the central role played by the time correlation functions. On the one hand, they are related by a Fourier transform to the response function; on the other hand, they facilitate useful expansions in terms of $\hbar$. The fact that the fermionic nature of the particles has been taken into account is an important aspect. In contrast to a recent attempt along similar lines [7] where complete integrability was assumed, we have here assumed fully developed chaos. Moreover, in the analysis of [7], the Berry–Tabor formula for periodic-orbit corrections has been used which fails for the interesting case of the harmonic oscillators. Since we now know that the Gutzwiller trace formula is exact for harmonic oscillators [2], and is also applicable for chaotic systems, we believe that our analysis may be extended to more general systems. As mentioned earlier, we draw attention of the reader to [30] where an expression for trace formula applicable to mixed phase space dynamics as occurring in the case of nearly integrable systems [32] is found.

There are many applications of such a formalism. In any discussion of relaxation phenomena, time correlation functions play an important role. To understand photoabsorption cross-section of atoms and molecules, semiclassical treatment along these lines has been used [1, 28]. In metallic clusters and quantum dots, there is a growing interest in the semiclassical treatment of the absorption spectra and cross-sections. The understanding of the plasmon modes in different metallic clusters, their splitting etc. are believed to have a semiclassical interpretation. However, for a systematic treatment of the collective modes, we need to consider two-body interactions, thus a Hamiltonian,

$$\hat{H} = \sum_i \hat{h}_i + \lambda \sum_{ij} \hat{v}_{ij}, \tag{51}$$

where $\hat{v}_{ij}$ denotes the two-body term. What we need to do here is a perturbative treatment of the two-body interaction by assuming that $\lambda \ll 1$ which, in turn, would give rise to collective modes.

A comment is here in order that the classical chaos, which may exist in the one-body effective potential obtained by the Hartree-Fock approximation, is only a property emerging out of the wave mechanics in the classical limit where the wavelengths are arbitrarily small. In this work, however, we use the classical periodic orbits in order to derive useful expansions of the quantum-mechanical quantities, which apply in the semiclassical regimes, i.e. before the classical limit is reached. The situation is similar as in other perturbative methods where a property valid for the (unperturbed) limiting system does not necessarily extend to the (perturbed) typical systems, and this without jeopardizing the validity of the perturbation method.

As a last comment, let us mention that, in a random-matrix framework, one can also study the correlation functions, and it has been recently shown that the time correlations depend upon the co-dimension of level crossing [33]. A comprehensive understanding of all the inter-relations will be one of the aims we look forward to.

Acknowledgements

SRJ is financially supported by the "Communaute Française de Belgique" under the ARC contract No. 93/98-166. SRJ is on an extraordinary leave from the Theoretical Physics Division, Bhabha Atomic Research Centre, Bombay 400085, India. PG thanks the National Fund for Scientific Research (F.N.R.S. Belgium) for financial support.

References

[1] P Gaspard, D Alonso and I Burghardt, *Adv. Chem. Phys.* **XC**, 105 (1995)
[2] M Brack and S R Jain, *Phys. Rev.* **A51**, 3462 (1995)
[3] S R Jain and S V Lawande, *Proc. Ind. Natl. Sci. Acad.* **A61**, 275 (1995)
[4] P van Ede van der Pals and P Gaspard, *Phys. Rev.* **E49**, 79 (1994)
[5] J A Folk, S R Patel, S F Godijn, A G Huibers, S M Cronenwett, C M Marcus, K Campman and A C Gossard, *Phys. Rev. Lett.* **76**, 1699 (1996)

[7] A Dellafiore, F Matera and D M Brink, *Phys. Rev.* **A51**, 914 (1995)

[8] S R Jain, A K Jain and Z Ahmed, *Phys. Lett.* **B370**, 1 (1996)

[9] D M Brink, A Dellafiore and M Di Toro, *Nucl. Phys.* **A456**, 205 (1986)

[10] H Nishioka, K Hansen and B R Mottelson, *Phys. Rev.* **B42**, 9377 (1990)

[11] V Strutinsky, A G Magner, S R Ofengenden and T Dossing, *Z. Phys.* **A283**, 269 (1977)

[12] M C Gutzwiller, *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1991)

[13] R Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (John Wiley and Sons, Inc., New York, 1974)

[14] A L Fetter and J D Walecka, *Quantum Theory of Many-Particle Systems* (McGraw Hill Inc., New York, 1971)

[15] B Eckhardt, S Fishman, K Müller and D Wintgen, *Phys. Rev.* **A45**, 3531 (1992)

[16] This famous trick is first known to have been applied by W Pauli (cf. ref. [14]). The relevant reference for the present context is the paper by P Gaspard and S A Rice, *Phys. Rev.* **A48**, 54 (1993)

[17] R Artuso, E Aurell and P Cvitanović, *Nonlinearity* **3**, 325, 361 (1990)
P Cvitanović and B Eckhardt, *J. Phys.* **A24**, L237 (1991)

[18] P Gaspard and D Alonso Ramirez, *Phys. Rev.* **A45**, 8383 (1992)

[19] Y Colin de Verdière, *Compos. Math.* **27**, 83, 159 (1973)

[20] J Chazarain, *Invent. Math.* **24**, 65 (1974); *Comm. Part. Diff. Eq.* **5** (6), 595 (1980)

[21] V P Maslov and M V Fedoriuk, *Semiclassical Approximation in Quantum Mechanics* (Reidel, Boston MA, 1981)

[22] A Voros, *Suppl. Prog. Theor. Phys.* **116**, 17 (1994)

[23] R Balian and C Bloch, *Ann. Phys.* (*New York*) **85**, 514 (1974)

[24] R Aurich, M Sieber and F Steiner, *Phys. Rev. Lett.* **61**, 483 (1988)

[25] M V Berry and J P Keating, *Proc. R. Soc. London* **A437**, 151 (1992)

[26] M L Du and J B Delos, *Phys. Rev.* **A38**, 1896, 1913 (1988)

[27] N L Balasz and B K Jennings, *Phys. Rep.* **104**, 347 (1984)

[28] D Alonso Ramirez, *Semiclassical Quantization and Classically Chaotic Systems*, Ph.D. Thesis (Université Libre de Bruxelles, unpublished, 1995)

[29] M V Berry and M Tabor, *Proc. R. Soc. London* **A349**, 101 (1976)
M V Berry and M Tabor, *J. Phys.* **A10**, 371 (1977)

[30] S C Creagh, *Ann. Phys.* **248**, 60 (1996)

[31] M Pollicott, *Invent. Math.* **81**, 413 (1985)
D Ruelle, *Phys. Rev. Lett.* **56**, 405 (1986)
J. Stat. Phys. **44**, 281 (1986); *J. Diff. Geom.* **25**, 99, 117 (1987)

[32] One can work with action-angle variables for this class of systems which helps in generalizing [29]. A systematic treatment of resonant periodic orbits was initiated by A M Ozorio de Almeida and J H Hannay, *J. Phys.* **A20**, 5873 (1987)

[33] P Gaspard, S R Jain and P van Ede van der Pals, *Time Correlation Functions of Complex Quantum Systems* (preprint, 1996)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 2,
February 1997
pp. 517–535

Accuracy of trace formulas

ARUL LAKSHMINARAYAN
Physical Research Laboratory, Navarangapura, Ahmedabad 380 009, India

Abstract. Using quantum maps, we study the accuracy of semiclassical trace formulas. The role of chaos in improving the semiclassical accuracy in some systems is demonstrated quantitatively. However, our study of the standard map cautions that this may not be most general. While studying a sawtooth map we demonstrate the rather remarkable fact that at the level of the time one trace even in the presence of fixed points on singularities the trace formula may be exact, and in any case has no logarithmic divergences observed for the quantum bakers map. As a byproduct we introduce fantastic periodic curves akin to curlicues.

Keywords. Quantum chaos; quantum maps; Gutzwiller's periodic orbit sum.

PACS No. 03.65

1. Introduction

The trace formula of Gutzwiller [1] provides a window into the semiclassical treatment of chaotic systems and is hence of considerable importance in many areas of physics and chemistry. Yet the understanding of this formula has been hampered by the complexity of the classical task involved, and often the semiclassical effort is greater than that for a full quantum solution which makes the semiclassical methods either only of academic interest or as tools for explaining features of a full quantum solution. In this situation simple models have played a key role just as has been the case with the understanding of classical chaos, as evidenced for example by the study of the logistic map and the smale horseshoe.

Although the trace formula (and its relatives) are important, the accuracy of these have not been rigorously established. The trace formula being the first term in an asymptotic expansion, has many approximations with apparently uncontrolled errors. The very many applications of the stationary phase approximation, on functional and on ordinary integrals obscures the nature of the errors. One ongoing issue is the *time* at which we can expect the various semiclassical approximations to go bad with a candidate the so called log time which is of the order of $-\log(\hbar)/\lambda$, applicable for chaotic systems with a Lyapunov exponent λ, being called too pessimistic [2]. We note that although the semiclassical formula for the time evolution (due to Van Vleck) has been essentially known almost since the beginning of quantum mechanics the nature of the errors is still a matter of interest, especially since the trace formula depends on a Laplace transform of the Van Vleck approximation [1].

The quantum bakers map has been studied as a model of quantum chaos as indeed it divergences. In fact we will show that in certain circumstances the semiclassical trace can be exactly the quantum trace.

In pursuance of certain results in [5], we use the existing models of quantum chaos to study the accuracy of the simplest trace of the quantum problem. To this end we will in particular use the much studied quantum standard map and the much less studied quantum sawtooth map both in the settings of the phase space as a cylinder and as a torus. These being maps derived from time periodic systems we study simply the trace of the relevant unitary Floquet operators. In § 2 the cylindrical phase space is used, while in § 3 the toral phase space is used, we end with some comments in § 4. In the Appendix we provide a self contained derivation of the "trace formula" for a class of maps on the torus that includes the maps studied below and all the so called kicked systems. While this is in principle known, it may be useful to recollect it here, as the other published proof of Tabor [17] is for the case of maps on the entire plane, and is hence quite different. Those with little exposure to the trace formula may peruse the Appendix first.

2. The cylindrical phase space

Consider the well known standard map $(q_{n+1} = q_n + Tp_n, p_{n+1} = p_n + K\sin(q_{n+1}))$ defined on the cylinder $[0, 2\pi) \times (-\infty, \infty)$ with periodic boundary conditions in q. The first iterative equation is then to be evaluated with a modulo 2π rule. A simple change of variables from p to Tp implies that the relevant parameter is just one in number and is $k = KT$. This area preserving map and its quantization have been studied in extensive detail as a model of chaos and its quantum manifestations and led to such important discoveries as the quantum suppression of chaotic diffusion. Some of these are reviewed in [7]. However its use in understanding the Gutzwiller trace formula has been limited. The work of Tabor [17] where he derived the relevant trace formula for quantum maps on the plane did use the standard map but no rigorous or extensive study has yet been attempted. The quantization of the standard map is described by an infinite dimensional unitary matrix with matrix elements in the momentum representation being

$$U_{nm} = (-i)^{n-m} J_{n-m}(k/T\hbar)\exp(-im^2\hbar T/2). \qquad (1)$$

Here $J_\nu(z)$ is a Bessel function. As has been noted earlier there are two relevant parameters quantum mechanically and we can take $T = 1$ while we do not have the freedom to choose $\hbar = 1$ anymore. The trace of the quantum map is then given by

$$\mathrm{Tr}(U) = J_0(k/\hbar) \sum_{n=-\infty}^{\infty} \mathrm{e}^{-i\hbar n^2/2} = J_0(k/\hbar)\Theta_3(0| - \hbar/2\pi), \qquad (2)$$

where Θ_3 is a Jacobi theta function. An application of the Poisson formula enables us to

Strictly speaking this sum need not exist and may without further regularization be infinite. This is not due to any peculiarities of maps, and in fact this is the rule in quantum mechanics, the trace of the propagator could be infinite in real time. However this will not worry us further as the semiclassical trace will also diverge exactly, allowing us to make meaningful relative error estimates.

Now we turn to a semiclassical evaluation of the quantum trace. We wish to then write the sum in terms of the period one or fixed points of the classical map. The Lagrangian generating the standard map on the cylinder is

$$L(q_n, q_{n+1}, l_n) = (q_{n+1} - q_n + 2\pi l_n)^2/2 - k\cos(q_{n+1}) \tag{4}$$

as $\partial L/\partial q_n = -p_n$ and $\partial L/\partial q_{n+1} = p_{n+1}$. l_n is the winding number in the q direction or $2\pi l_n$ is subtracted by the modulo operation performed at time n to keep the map on the cylinder. The fixed points are determined by the condition $(q_{n+1} = q_n, p_{n+1} = p_n)$, and are therefore simply given by $(q = 0, p = 2\pi l)$ and $(q = \pi, p = 2\pi l)$, where l runs over the integers. We note that the number of fixed points are infinite and that their positions do not depend on the classical chaos parameter k. The denominator in the Gutzwiller trace formula (or its versions for the case of quantum maps, they are essentially same) has the factor $\sqrt{\mathrm{Tr}(J) - 2}$, where J is the classical stability matrix associated with the periodic orbit. In the case of standard map fixed points this is either $\sqrt{k}$ or $\sqrt{-k}$ depending on whether $q = 0$ or $q = \pi$.

The semiclassical approximation is of the form

$$\mathrm{Tr}(U)_{\mathrm{sc}} = \sum_{\text{fixed points}} \exp(iS/\hbar)/\sqrt{(\mathrm{Tr}\,(J) - 2)}, \tag{5}$$

where S is the classical action and since we are dealing with discrete time systems with the time interval being taken as unity this is identical to the Lagrangian written above. We get for the standard map on the cylinder the following semiclassical approximation for the trace (see Appendix):

$$\mathrm{Tr}(U)_{\mathrm{sc}} = \sqrt{\frac{4}{ik}}\cos\left(\frac{k}{\hbar} - \frac{\pi}{4}\right) \sum_{l=-\infty}^{\infty} \exp(2\pi^2 i l^2/\hbar). \tag{6}$$

This then has to be compared with (3). The asymptotic form of $J_0(x)$ for large arguments is [8]

$$J_0(x) \sim \sqrt{\frac{2}{\pi x}}\left[\cos\left(x - \frac{\pi}{4}\right)P(x) - \sin\left(x - \frac{\pi}{4}\right)Q(x)\right],$$

where

$$P(x) = 1 - 9/(128x^2) + \cdots \quad \text{and} \quad Q(x) = -1/(8x) + \cdots,$$

where the neglected terms are at most of order x^{-}. We then derive that

$$\frac{\mathrm{Tr}(U)}{\mathrm{Tr}(U)_{\mathrm{sc}}} = P\left(\frac{k}{\hbar}\right) - \tan\left(\frac{k}{\hbar} - \frac{\pi}{4}\right) Q\left(\frac{k}{\hbar}\right). \tag{7}$$

Even this simple result is instructive. It indicates spectacular failures of the trace formula when the ratio $k/\hbar$ passes through $\pi/4 + (2m+1)\pi/2$, where m is a positive integer. At these points the semiclassical trace becomes zero and the exact quantum trace is also small indicating that such a failure will not be serious for the evaluation of the spectra. But it does suggest that the ratio of the quantum to the semiclassical need not be bounded in the entire range of the parameters. This divergence of the ratio is thus not due to any coalescing stationary points or bifurcations in general; as we have already noted the fixed points are independent of the chaos parameter k. Second, it is clear that for a larger k or chaos the semiclassical trace is going to be a better approximation. The authors in [2] have observed this, and we see that the simplest quantum trace provides a quantitative evidence of this. To compare the absolute error rather than the relative, we have to deal with the Jacobi theta function which occurs identically both in the quantum and in the semiclassical traces, and as this can depend critically on the rationality of $\pi/\hbar$ we do not pursue this here, except note that in unbounded phase spaces such as the plane and the cylinder the absolute value of the traces might actually diverge.

The case of the standard map has been extensively studied but not so the quantum mechanics of the sawtooth map, which is in many ways simpler because it is a piecewise linear map for which it is easy to extract classical information. Consider the map $(q_{n+1} = q_n + Tp_n, p_{n+1} = p_n + K(q_{n+1} - \pi))$ defined on the cylinder $[0, 2\pi) \times (-\infty, \infty)$, which is completely chaotic for positive K. This can be considered as a simple harmonically kicked particle in one dimension. We may view the system as an otherwise free particle on a unit circle being subjected to a kick whose strength is proportional to the physical angle and which has a single discontinuity when the angle is zero. The trace of the quantum propagator may be written down just as in the case of the standard map $(T = 1)$with the result that

$$\mathrm{Tr}(U) = \frac{1}{\sqrt{K}} \,\mathrm{Erf}\left(\mathrm{e}^{-i\pi/4}\sqrt{\frac{K}{2\hbar}}\right) \sum_{n=-\infty}^{\infty} \exp(2\pi^2 in^2/\hbar), \tag{8}$$

where Erf is the error function.

The semiclassical trace is also found quite easily, because the fixed points are at $(q = \pi, p = 2\pi l)$, l being any integer. The classical action is similar to that of the standard map with the potential term being replaced by a periodicised quadratic function. Thus we get

$$\mathrm{Tr}(U)_{\mathrm{sc}} = \frac{1}{\sqrt{K}} \sum_{n=-\infty}^{\infty} \exp(2\pi^2 in^2/\hbar), \tag{9}$$

and the ratio of the quantum to semiclassical simply becomes

$$\frac{\mathrm{Tr}(U)}{\mathrm{Tr}(U)_{\mathrm{sc}}} = \mathrm{Erf}\left(\mathrm{e}^{-i\pi/4}\sqrt{\frac{K}{2\hbar}}\right). \tag{10}$$

Inserting the asymptotic form for the error function we get the relative error to be

$$\left|1-\frac{\mathrm{Tr}(U)}{\mathrm{Tr}(U)_{\mathrm{sc}}}\right| \approx \sqrt{\frac{2\hbar}{\pi K}}. \tag{11}$$

The neglected terms are at most of order $(\hbar/K)^{3/2}$.

In this map also we see the seemingly general result that the larger the chaos (larger K) the semiclassical accuracy is better. The precise way in which this occurs for this map is like $K^{-1/2}$. The difference between this map and the standard map is that the semiclassical approximation seems more well behaved and the sudden decline in the accuracy in the case of the standard map, due to the mismatch between the zeros of the quantum and the semiclassical is absent here. However in the regions where the semiclassical (and quantum) traces are not very close to zero, the standard map would perform better as the errors are of order $\hbar/k$ while for the sawtooth map the errors are of order $(\hbar/k)^{1/2}$. We will see that surprisingly many of our observations will remain valid in the toral setting as well.

3. The toral phase space

Two dimensional maps on the torus (periodic boundary conditions in *both* position and momentum) have been studied extensively, from the standard map to the cat map of Arnold. Their quantizations have also been investigated by many [3, 7, 9], as this is a legitimate model for a bound Hamiltonian system with two degrees of freedom, the minimum needed for chaos. Unlike the case of the cylindrical phase space where scattering may take place, sheer topology now prohibits this on the torus. The act of taking periodic boundary conditions on the momenta is reminiscent of a similar procedure in condensed matter physics which has also provided an interesting two dimensional map, namely the kicked Harper model. Quantum mechanics now is on a finite dimensional Hilbert space and we take the torus to be a unit torus (the dissected square is of unit sides). An alternate view of toral quantum maps is as the case of "resonance" of the cylindrical map in which case the quantum mechanics can be exactly reduced to that of a torus. Position and momentum take discrete values which we take to be n/N, where N is the dimension of the Hilbert space and $n = 0, \dots, N-1$. Here N is required to be an *even integer* and is the inverse of the Planck constant. Therefore the classical limit is $N \to \infty$.

In this section we will first deal with the sawtooth map on the torus. We will write the map a little differently for convenience as $(q_{n+1} = (q_n + p_n)\mathrm{mod}1,\ q_{n+1} = (p_n + Kq_{n+1})\mathrm{mod}1)$. The quantum map or propagator is given by

$$U_{nm} = \frac{e^{-i\pi/4}}{\sqrt{N}} \exp(i\pi K m^2/N) \exp(i\pi(n-m)^2/N). \tag{12}$$

This then is the quantum equivalent of the classical map restricted on a torus. It is a finite unitary matrix as all quantum maps on compact phase spaces are. Clearly it is one of the simplest quantum objects, having elementary functions as matrix elements. Yet we find that it is rich enough to reflect many essential features of the effect of classical chaos on quantum systems. For more properties of a very similar quantum map please refer [10].

The classical Lagrangian (or action) for this map is

$$L(q_n, q_{n+1}, l_n, m_n) = (q_{n+1} - q_n + l_n)^2/2 + K(q_{n+1}(\mathrm{mod} 1))^2/2 - m_n q_{n+1}, \tag{13}$$

where the integers l_n and m_n are winding numbers in the q and p directions respectively. The fixed points are given by $(q = m/K, p = 0)$ where $m = 0, 1, 2, \ldots, [K]$. The fixed point at the origin is in fact on the line of discontinuity, and is quite like the bakers map and its quantization. In that case it was noted that such singular periodic orbits, "half hyperbolic ", could give rise to unusual semiclassical behaviour and that the quantum and semiclassical could diverge semiclassically as $\log(\hbar^{-1})$. A bakers map was constructed to explicitly avoid this situation and a recovery of standard expectations was realized [6]. However if we take the above quantization as valid (since it follows canonical quantization closely [11]) it will turn out that the fixed point on the discontinuity is benign in the case of the sawtooth map, in that it does not lead to semiclassically divergent terms. This may suggest that the unusual nature of the bakers map originates either from the unusualness of the very quantization procedure adopted or is peculiar to the bakers map.

Let us look at this case more closely, and take $0 < K < 1$, so that there is a single fixed point at the origin which is also a point of discontinuity. A direct application of the Gutzwiller formula will be incorrect in this case, but the rectification is obvious. The stationary phase integral must not extend over the entire real line, but must be restricted to $[0, \infty)$. Thus the semiclassical approximation is then the approximation (quantum $\sim$ semiclassical)

$$\frac{e^{-i\pi/4}}{\sqrt{N}} \sum_{n=0}^{N-1} \exp(i\pi K n^2/N) \sim \frac{1}{2\sqrt{K}}. \tag{14}$$

An estimate due to Wilton which we give below bounds the difference as $N \to \infty$, but this is not a good estimate of the errors (for $0 < K < 1$), as these are of the order of the sums themselves. An application of the Euler–Maclaurin summation rule will give us the above approximation quite simply. But more important is the observation that it is legitimate to replace summations with integrals even in the presence of fixed points sitting on discontinuities, a rule that was flagrantly violated in the case of the bakers map.

Let the errors be $\epsilon(K, N) = |\mathrm{Tr}(U) - \mathrm{Tr}(U)_{\mathrm{sc}}|$, that is the absolute value of the difference between the right hand and the left hand side of the (14). We now observe some remarkable number theoretic properties of these errors. First we note the surprising fact that if N is a multiple of 4, $\epsilon(1/2, N) = 0$. In this case not only does the semiclassical and quantum converge in the classical limit, they are exactly the same (at the level of the trace), despite the presence of the fixed point on the discontinuity! We have noted before that the main feature which makes the sawtooth map generic, in the sense that the semiclassical is only an approximation, and in the sense of eigenfunctions scarring and spectral random matrix properties is that the incomplete Gauss sums that arise while compounding propagators do not reduce to simple forms [10], unlike in the case of the cat maps (K an integer) where we get complete Gauss sums and has been investigated by Hannay and Berry [9] and the extensive number theoretic properties were explored by

Keating [12]. Thus it is a bit interesting that an incomplete Gauss sum can be done exactly and is the semiclassical approximation.

Consider the sum

$$\sigma_N = \sum_{n=0}^{N-1} \exp(i\pi n^2/2N), \tag{15}$$

which is incomplete because the period of $s_n = \exp(i\pi n^2/2N)$ is $2N$ and the sum is only over the first N terms. First we note that $s_{N/2+m} = s_{N/2-m}\exp(i\pi m)$, thus terms with odd n do not matter as they cancel out of the sum exactly. Let

$$\Sigma_{2N} = \sum_{n=0}^{2N-1} s_n$$

which is a complete Gauss sum [9]. But also note that if N is a multiple of 4,

$$s_{n+N} = (-1)^n s_n.$$

This then implies that

$$\sigma_N = \tfrac{1}{2}\Sigma_{2N} = \tfrac{1}{2}\sqrt{2N}\,\mathrm{e}^{i\pi/4},$$

therefore

$$\mathrm{Tr}(U) = \mathrm{e}^{-i\pi/4}\sigma_N/\sqrt{N} = 1/\sqrt{2}. \tag{16}$$

If N is not divisible by 4, then $\epsilon(1/2, N)$ is no longer zero and the semiclassical result is an approximation. Thus there are two distinct classes when $K = 1/2$ as far as the behaviour of the error is concerned, This can be generalized and we observe that if $K = p/q$, is a rational number, with relatively prime integers p and q, then there are as many classes as there are possible remainders of even N, on division by q. Within each class the error monotonically goes to zero in the classical limit, $N \to \infty$. We have observed numerically that for large N to an excellent approximation

$$\epsilon(K,N) = C(K)N^{-1/2}. \tag{17}$$

Figure 1(a) shows the case $K = 2/5$, when there are five classes. The constant $C(K)$ depends on both K and the class, but we have observed that $C(K) = C(1-K)$. If K is irrational, the number of classes is infinite, there is no monotonic decrease of error, but the general trend is to decrease, as is evident from figure 1(b), where we have taken K to be the most irrational number $\gamma = (\sqrt{5}-1)/2$, the golden mean. The number theoretical properties of the error is of course a reflection of the sum itself, and a fuller study is thus desirable.

We however shall now turn to the case when K is larger than one, and there are a number of fixed points. If K is much larger than unity we need not correct for the corner fixed point as the errors incurred will be of the order $K^{-1/2}$, which is anyway the order of the total error according to Hardy, Littlewood and Wilton, as quoted in [13]. The semiclassical approximation is now described by

$$\frac{\mathrm{e}^{-i\pi/4}}{\sqrt{N}} \sum_{n=0}^{N-1} \exp(i\pi K n^2/N) \sim \frac{1}{\sqrt{K}} \sum_{n=0}^{[K]} \exp(-i\pi N n^2/K). \tag{18}$$

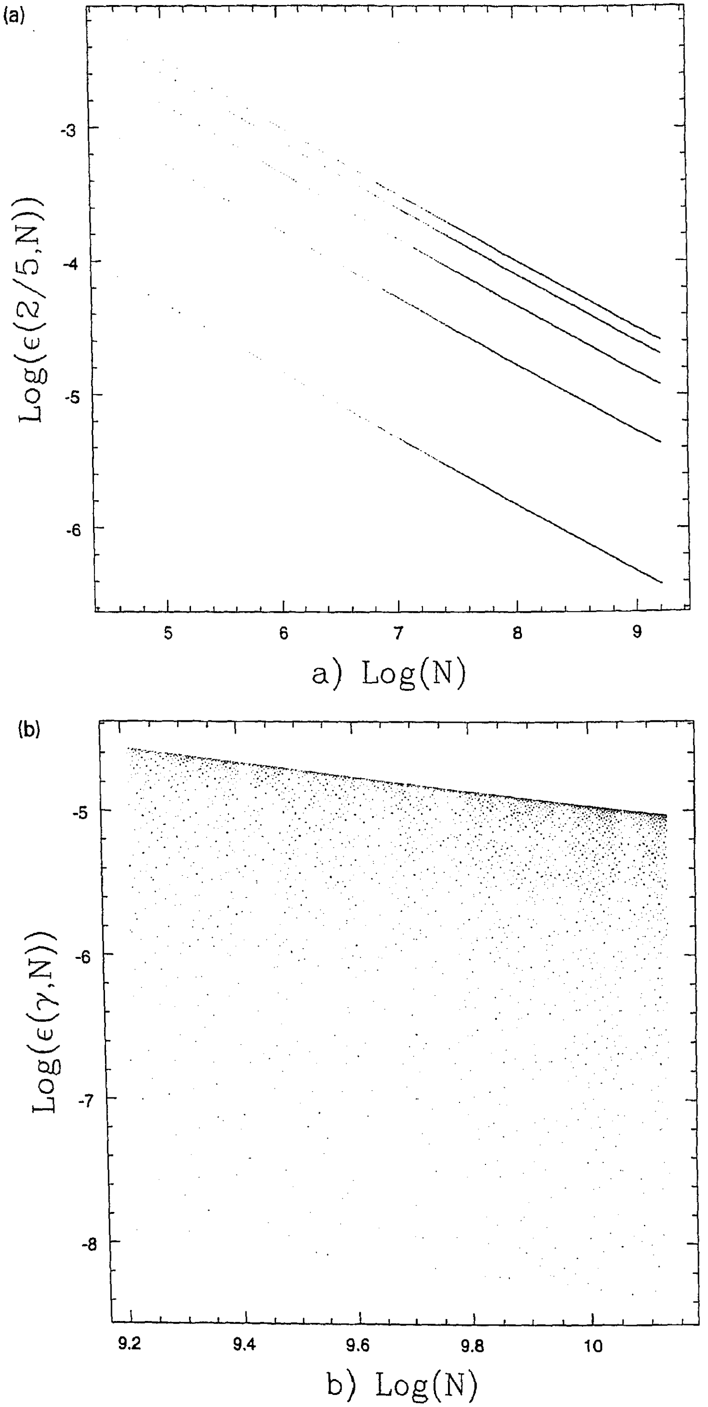


Figure 1. The error as a function of inverse Planck's constant N, for the sawtooth map with (**a**) $K = 2/5$ (**b**) $K = \gamma$.

Here $[K]$ is the integer value of K. Note that this approximation clearly demonstrates the conjugacy between the Planck constant $(1/N)$ and the chaos parameter K. It is a good approximation if K is fixed and N tends to infinity, namely the classical limit, but also in the opposite limit of extreme "chaos", and small quantum system. For example when K tends to infinity for $N = 2$, a two level system. Undoubtedly, this approach to two level systems is to say the least inadvisable, as the quantum problem is much easier to solve. Nevertheless it throws up some interesting questions, because what would one mean by extreme chaos in a two level system? Quantum chaos for legitimacy in using the epithet chaos has always relied on the classical limit, and in the extreme quantum limit the concept becomes nebulous.

The above approximation is derived from the Gutzwiller formula, using the $[K] + 1$ fixed point actions and the stability matrix. The fixed point at the origin is included wrongly in the formula and we have discussed it separately in the case $0 < K < 1$ above. Now we use an estimate of Wilton, quoted in [13], to bound the error in this approximation. This estimate is the inequality

$$|\mathrm{Tr}(U) - \mathrm{Tr}(U)_{\mathrm{sc}}| < \frac{a}{\sqrt{N}} + \frac{b}{\sqrt{K}}, \tag{19}$$

where a and b are constants of order one whose values are slightly different from the quoted ones, as the sawtooth sum is from 0 to $N - 1$, rather than 1 to N, but this is a trivial complication. The essential conclusion is that the error is bounded in N, and corroborates our conclusion that there is no $\log(N)$ problems in the sawtooth map

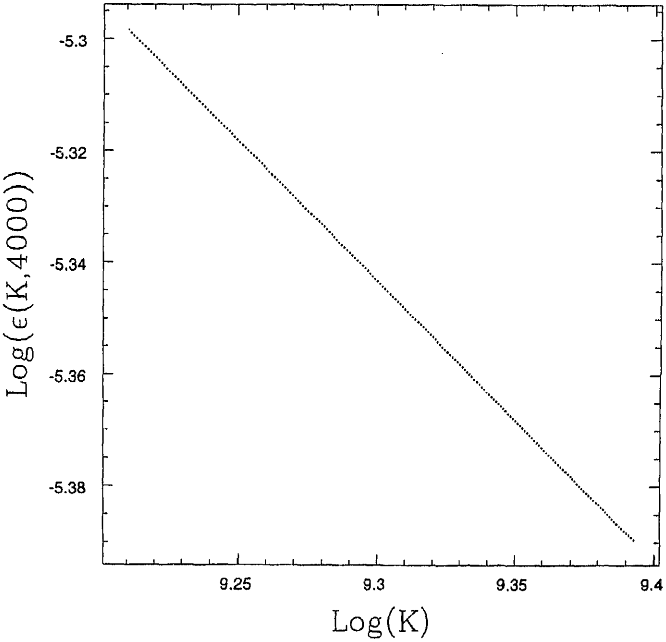


Figure 2. The error as a function of the chaos parameter K for the sawtooth map with $N = 4000$.

despite the presence of the fixed point on the singularity. Also we once again notice that a larger K (larger chaos) implies a better semiclassical accuracy. In figure 2 we show the results of a numerical computation of the error, which shows that the error goes roughly as $K^{-1/2}$, with the slope depending weakly on N and K and rather strongly on the fractional part of K. Notice that at integer K (cat maps) the approximation is exact. We thus note that like its relative on the cylinder the torus map error is also of the order of $K^{-1/2}$.

We had earlier studied the sawtooth map on the torus with slightly different boundary conditions [10], and came to essentially similar conclusions, however in that case there was no fixed point on the discontinuity. When the fixed points moved towards the singularity, at the exact point of contact, the map turned into the continuous cat map when the trace formula becomes exact. We now turn to a smooth map on the two torus, the standard map. Much studied in other contexts, it is largely unstudied from the view of trace formulas, although it should form a natural object of study, as it represents generic mixed systems, and one recognizes the difficulties of such systems apropos trace formulas.

We will consider the map once again on the unit torus. The map is now $(q_{n+1} = (q_n + p_n)\mathrm{mod}1, p_{n+1} = (p_n + k\sin(2\pi q_{n+1}))\mathrm{mod}1)$. The Lagrangian (action) is

$$L(q_n, q_{n+1}, l_n, m_n) = (q_{n+1} - q_n + l_n)^2/2 - k\cos(2\pi q_{n+1})/2\pi - m_n q_{n+1}. \qquad (20)$$

The fixed points are easy to find and are at $(q_m = (1/2\pi)\arcsin(m/k), p_m = 0)$, where m is an integer such that $|m| < k$. Once again the location and number of the fixed points are very much dependent on the the chaos parameter k unlike the standard map on the cylinder. Note that there are two fixed points associated with each m.

The quantum propagator is again given by (N even throughout)

$$U_{nm} = \frac{e^{-i\pi/4}}{\sqrt{N}} \exp(-ikN\cos(2\pi n/N))\exp(i\pi(n-m)^2/N). \qquad (21)$$

and the trace is the finite sum

$$\mathrm{Tr}(U) = \frac{e^{-i\pi/4}}{\sqrt{N}} \sum_{n=0}^{N-1} \exp(-ikN\cos(2\pi n/N)). \qquad (22)$$

The semiclassical approximation must approximate this sum using the classical fixed points, their actions and their stabilities. First note that the quantum trace apart from the factor $e^{-i\pi/4}$ is real. The sawtooth map traces were Gaussian sums, and partial sums of these give rise to the "curlicues", that have been studied in detail earlier and find applications from optics to quantum mechanics [13]. The partial sums associated with the quantum trace of the standard map ought to be interesting mathematical objects in their own right.

Here we present a preview of the intriguing class of periodic "curves" that this gives rise to. Consider the sum

$$\sigma_N(k,l) = \sum_{n=0}^{l} \exp(-ikN\cos(2\pi n/N)). \qquad (23)$$

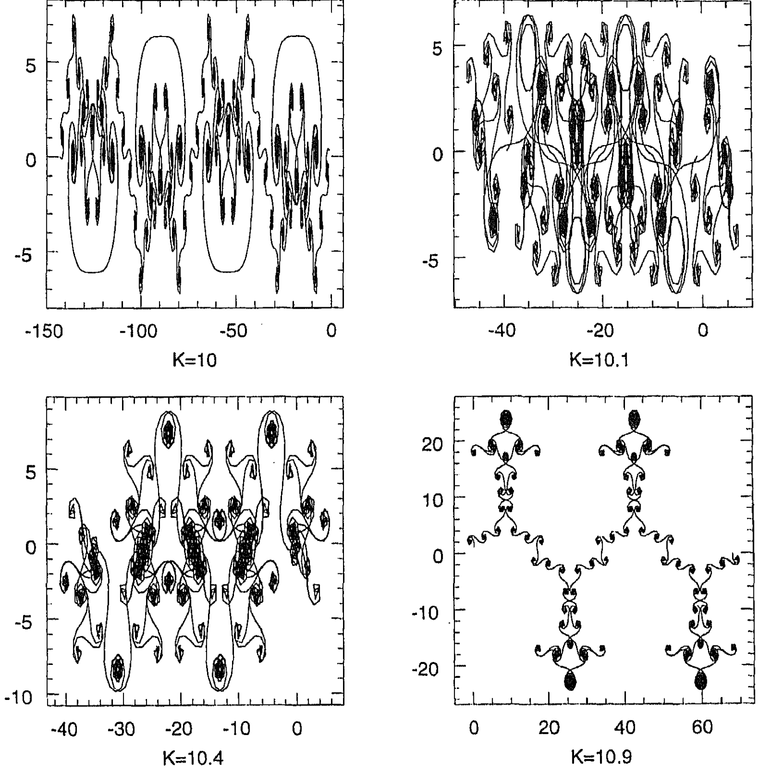


Figure 3. A panel of the partial sums $\sigma_{1000}(K, l)$ (23) for various K.

The quantum standard map trace is essentially $\sigma_N(k, N)$. We plot in figure 3 the real, $x(l)$, and imaginary, $y(l)$, parts of $\sigma_N(k, l)$ as an argand diagram. Strictly there are only as many points as terms in the partial sum, and as such do not define a curve, but the connections showing the motion of the points in "time" l is suggestive. Only the sum up to $l = N/2$ is relevant as beyond this a reflection and periodic translation is the whole curve. Hence unlike curlicues, these are periodic curves. A plethora of cornu spirals interweave to conjure the extravagantly ornate features. Several "eyes" of cornu like spirals are connected in a remarkable manner. An essential feature of these curves is their extreme susceptibility to small changes in k, especially at large N values. We cannot rule out the possibility of number theoretic features here as well, and indeed we observe marked difference between the cases when k is an integer and when it is not.

The periodic nature of the curves are easily understood, as $y(l + N/2) = -y(l)$ and $x(l + N/2) = x(l) + A$, where $A = \sum_{n=0,N/2-1} \cos(-ikN\cos(2\pi n/N))$. Hence the period of the curves are $2A$. As N is increased for fixed k the curves tend to smoothen out more but with the cornu spiral structures intact. Also note that the quadratic approximation of the cosine term in the exponential are the culicues studied earlier [13].

A detailed study of these curves is however not pursued in this paper, and we turn once more to the central theme, namely the accuracy of semiclassical approximations. Define

$$\theta_N(l) = N(\sqrt{k^2 - l^2} + l \arcsin(l/k)). \tag{24}$$

Then a careful application of the trace formula yields the semiclassical approximation

$$\mathrm{Tr}(U)_{\mathrm{sc}} = \frac{\mathrm{e}^{-i\pi/4}}{\sqrt{2\pi}} \left[\frac{2\sin(Nk + \pi/4)}{\sqrt{k}} + \sum_{l=1}^{[k]} \frac{4\sin(\theta_N(l) + \pi/4)}{(k^2 - l^2)^{1/4}} \right], \tag{25}$$

which is real except for the factor $\mathrm{e}^{-i\pi/4}$, just as the exact quantum trace in (22). Here the fixed point at the origin has been treated separately as corresponding to each $|m|$ there are four fixed points except when $m = 0$, which gives rise to two fixed points. The range of the inverse sine function is $[0, \pi/2]$. Note that this approximation can be derived independent of the trace formula, by first converting the sum into an infinite sum of integrals via the Poisson summation formula, and of these integrals retaining only those which contribute to the stationary phase, and evaluating these integrals in the quadratic approximation. Indeed the trace formula in the context of maps on the torus is derivable by this procedure and is hence considerably simpler than the derivation for general Hamiltonian flows (see Appendix).

First we note a drawback of the semiclassical approximation. Near parameter values where bifurcations are going to create new periodic points the semiclassical

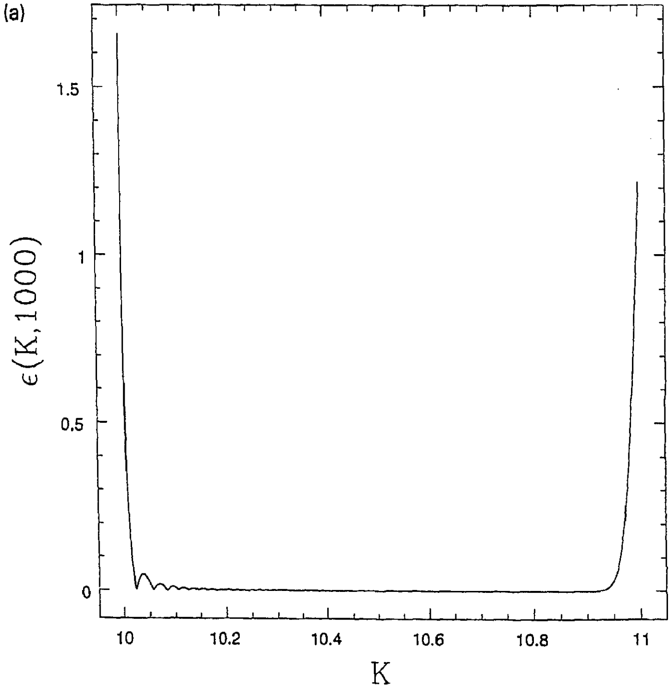


Figure 4a. (*Continued*)

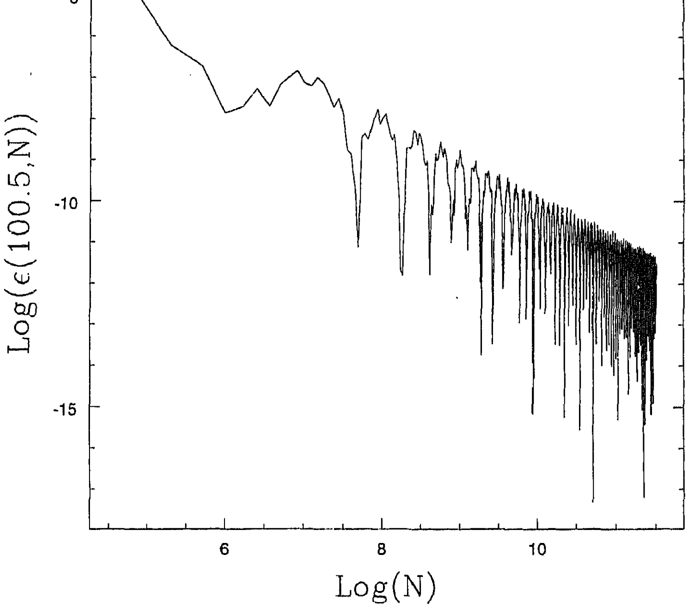


Figure 4. The error for the standard map on the torus as a function of (**a**) k, with $N = 1000$, (**b**) inverse Planck's constant N, with $k = 100.5$.

approximation becomes poor. Uniform approximations may be made to smooth over these parameter values. For instance in the case of the standard map, whenever k passes an integer value it creates four new fixed points via tangent bifurcations, and the approximation to the trace becomes poor. This is illustrated in figure 4(a). However notice that the approximation is indeed quite good even just away from these values. We will below retain the approximation made above and keep away from points of bifurcations to study various aspects of the error.

In figure 4(b) we show the absolute value of the error $\epsilon = |\mathrm{Tr}(U) - \mathrm{Tr}(U)_{sc}|$ for a fixed value of $k = 100.5$, and a range of N. It is clear that the errors involved are not monotonically decreasing with N, but that there are fluctuations in the mean. The mean however very nearly follows N^{-1} behaviour. This should be contrasted with the more cleaner but slower $N^{-1/2}$ behaviour of the sawtooth map. In general the errors incurred for the smooth standard map is smaller than those of the sawtooth map. We see that this is a property that is carried over from the maps on the cylinder. It is unclear to the author whether the standard map fairs better due to its smoothness or due to some other property. We note that the trace formula approximation behaves better with increasing chaos in the case of the sawtooth map than the standard map.

At a fixed value of N the behaviour of the trace for increasing k is quite complex. No simple rule like, higher chaos implies better semiclassical accuracy emerges. First we note that the ratios or the relative error fair quite poorly, with large abrupt deviations.

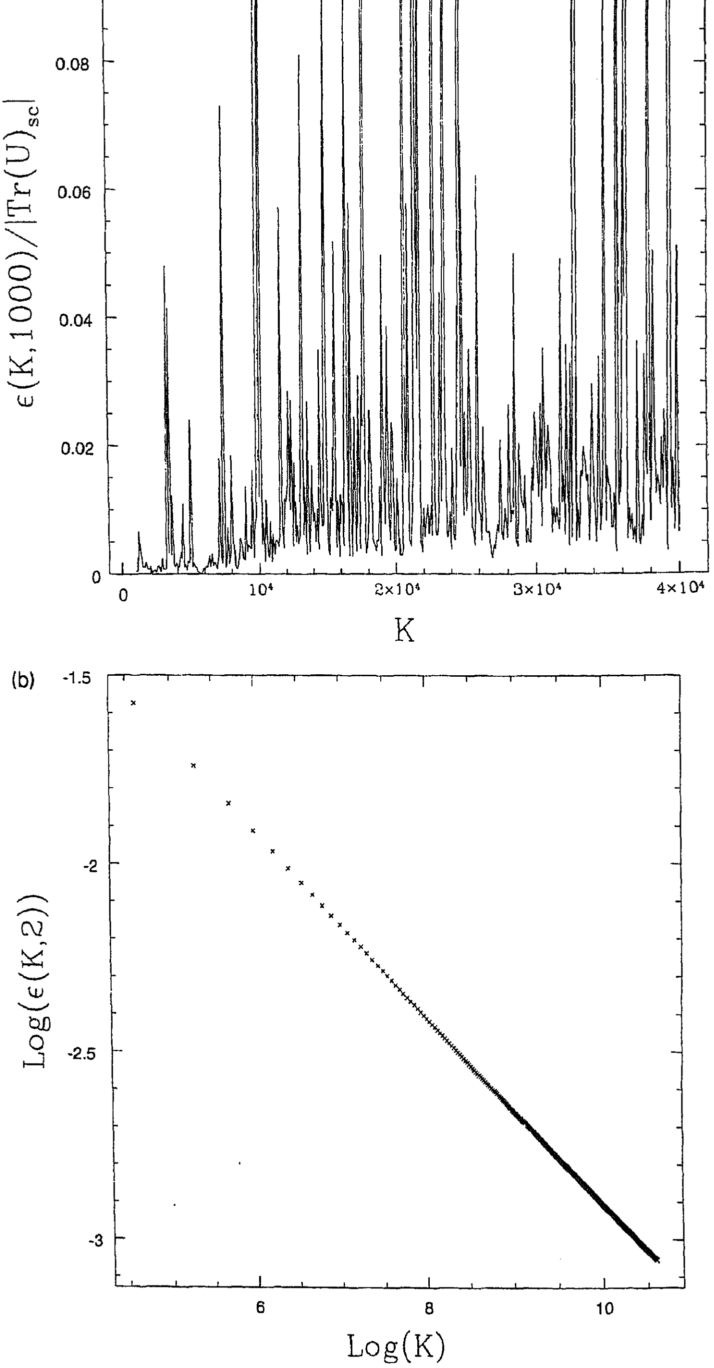


Figure 5. (**a**) The relative error for the standard map, as a function of the chaos parameter k, $N = 1000$, (**b**) the absolute error as a function of k in the extreme quantum limit, $N = 2$.

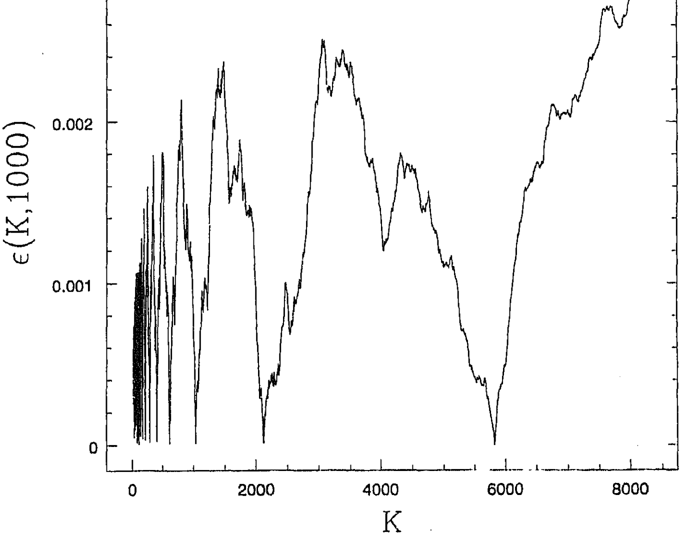


Figure 6. The error as function of k in a wide range of k, for the standard map with $N = 1000$.

Figure 5(a) shows the relative errors $|\epsilon|/|\mathrm{Tr}(U)_{sc}|$, but the range in the y axis has been lessened to highlight the predominant features. The error could become as high as 40 in the range of k shown. This is however easy to understand in the context of (7) for the map on the cylinder, as due to the mismatch of the zeros of the semiclassical and the quantum traces. We expect the absolute value of the errors to fair better. Indeed when N is small, one observes an increase of accuracy with k, and this is a power law for small N. Thus we show in figure 5(b) the logarithm of the errors with the logarithm of k, and we observe the rule that $\epsilon = Ck^{-\gamma}$, with $\gamma \sim 0.25$, for the extreme quantum case $N = 2$. Here the k go in steps of unity with their fractional part begin uniformly 0.5.

However as N is increased we observe that this is no longer true, and that the errors tend to oscillate with k in a fairly complex manner. The overall tendency seems to be in fact a deterioration of the accuracy with k. Figure 6 shows the errors as a function of k in a large range. The interesting aspect is that the oscillation period seems to increase geometrically. Therefore there is a possibility that the errors scale with k, but in a complex manner. However increasing k even higher seems to render the errors an almost monotonic growth, and we cannot comment on the many details until further analytic estimates are made and computer generated errors are under control.

An important difference between the semiclassics of the sawtooth map and the standard map is that because of the piecewise linear nature of the sawtooth map there was no need to truncate Taylor expansions about the periodic orbits, while in the case of the standard map this truncation is done up to second order. Higher order corrections can be worked out and have formed the subject of recent works [15, 16], it should be interesting to compare the errors when such higher order terms are included.

4. Summary

Using maps as models of quantum chaos we have studied the simplest traces from the point of view of semiclassical approximations, hoping that this will reflect some more general aspects of errors involved in the Gutzwiller's periodic orbit sum or trace formula. We found that indeed the maps on the cylinder had controllable relative errors. We found that the smooth standard map faired better than the sawtooth map in the order of magnitude of the errors. We also noted that the mismatch of zeros between the semiclassical and the quantum can lead to deterioration of relative errors. In the case of toral maps we displayed a rather remarkable fact that even in the presence of fixed points on discontinuities the trace formula may be exact at the level of the time one trace. Otherwise we showed that more generally there is no $\log(\hbar)$ divergences in the sawtooth map, unlike the case of the quantum bakers map. We also noted that the error may actually decrease with increasing chaos, a rather counter intuitive fact, which has quantitative evidence in the case of the sawtooth map. However we caution that this cannot be a general principle as a study of the standard map has shown a much more complex behaviour of the error. It is indeed a matter of importance to study and contrast the two maps, and understand why the errors behave qualitatively differently.

We have restricted our study to the time one trace and indeed it is natural and necessary that we study the higher traces. For instance to study the errors involved in the semiclassical evaluation of $\mathrm{Tr}(U^2)$ leads us into the problem of enumerating period two orbits, which can be fairly easily computed in the cases studied, but which are much larger in number than the fixed points. Preliminary results in this direction indicate that for the sawtooth map the error behaviour of the time two trace is similar to that of the trace, although the errors are larger. We have also displayed some remarkable curves akin to curlicues that arise out of a study of the trace of the standard map on the torus.

Appendix

Consider the following map defined on the unit torus:

$$\begin{aligned} q_{j+1} &= q_j + p_j - l_j \\ p_{j+1} &= p_j - V'(q_{j+1}) - m_j, \end{aligned} \tag{A1}$$

where $V(q)$ is a periodic "kicking" potential with unit periodicity, the prime indicates derivative, and the integers m_j and l_j are such that the map is restricted to the unit torus (winding numbers). They are the "modulo one" operations. As is well known the above map can be derived from a time dependent Hamiltonian system. Also there exists a generating function for the map on the torus given by

$$S(q_j, q_{j+1}; l_j, m_j) = \tfrac{1}{2}(q_{j+1} - q_j + l_j)^2 - V(q_{j+1}) - m_j q_{j+1} \tag{A2}$$

from which the map can be derived as $p_j = -\partial S/\partial q_j$ and $p_{j+1} = \partial S/\partial q_{j+1}$.

The quantum map corresponding to the (A1) is given by the finite dimensional unitary matrix

$$\langle n|U|n'\rangle = \frac{e^{-i\pi/4}}{\sqrt{N}} \exp\left(\frac{i\pi}{N}(n-n')^2 - 2\pi i N V\left(\frac{n}{N}\right)\right). \tag{A3}$$

This takes on the function of the propagator as it connects states after consecutive kicks. Here the representation is in the position basis and n and n' may take the integer values from 0 to $N-1$. N is itself the dimensionality of the Hilbert space and is also the inverse Planck constant, and is restricted to be even integers for preservation of toral boundary conditions. Semiclassics means the study of the unitary operator as $N \to \infty$.

The object of primary interest below is the trace of the powers of the propagator from which the spectrum may be derived by a Fourier transform. This can be written as

$$\mathrm{Tr}(U^T) = \frac{e^{-i\pi T/4}}{N^{T/2}} \sum_{n_i=0}^{N-1} \exp\left(\frac{2\pi i}{N}\sum_{j=1}^{T}\left(n_j^2 - n_j n_{j+1} - N^2 V\left(\frac{n_j}{N}\right)\right)\right), \tag{A4}$$

with

$$n_{T+1} = n_1.$$

It is implied in the above that the outer sum is over the T variables $n_i, i = 1, 2, \ldots, T$. Using the Poisson summation formula we can rewrite the above as

$$\mathrm{Tr}(U^T) = e^{-i\pi T/4} N^{T/2} \sum_{k_i=-\infty}^{\infty} \int_{-\epsilon}^{1-\epsilon} dx_1 \cdots dx_T$$
$$\exp\left(2\pi i N \sum_{j=1}^{T}\left(k_j x_j + x_j^2 - x_j x_{j+1} - V(x_j)\right)\right), \tag{A5}$$

with

$$x_{T+1} = x_1 \quad \text{and} \quad k_{T+1} = k_1.$$

The stationary phase approximation, assuming that N is large, gives the following T conditions

$$x_{j+1} = 2x_j - x_{j-1} - V'(x_j) + k_j; \quad j = 1, \ldots, T, \quad x_0 = x_T, \quad x_1 = x_{T+1}, \tag{A6}$$

with the further condition that

$$0 \le x_j < 1.$$

These are precisely the equations that determine the period T orbits of the map. The integers k_j are related to the integers l_j and m_j of (A1) by the relation $k_j = l_{j-1} - l_j - m_{j-1}$. With the assumption that they uniquely determine a periodic orbit of period an integer fraction of T, we restrict the infinite sum to only those that relate to such a periodic orbit labeled below by β. This is the first approximation and the second one is that we will extend the ranges of integration to the whole real line. These approximations together are exact in the case of the cat maps [9, 11, 12], and the third retaining up to the quadratic terms. This approximation is unnecessary for the piecewise linear sawtooth maps [10], but is necessary in general.

Therefore we define new variables y_j such that $x_j = q_j + y_j$ and q_j is the central periodic orbit.

$$\mathrm{Tr}(U^T) \sim \mathrm{e}^{-i\pi T/4} N^{T/2} \sum_{\beta} \exp\left(2\pi i N \sum_{j=1}^{T} (-q_j q_{j+1} + q_j^2 + q_j k_j - V(q_j))\right)$$

$$\times \int_{-\infty}^{\infty} \mathrm{d}y_1 \cdots \mathrm{d}y_T \exp\left(2\pi i N \sum_{j=1}^{T} (y_j^2 - y_j y_{j-1} - y_j^2 V''(q_j)/2)\right). \quad \text{(A7)}$$

The sum

$$\sum_{j=1}^{T} (-q_j q_{j+1} + q_j^2 + q_j k_j - V(q_j)) \quad \text{(A8)}$$

can be identified with the action of the periodic orbit, as from (A2) one gets

$$S_\beta = \sum_{j=1}^{T} (q_j^2 - q_j q_{j+1} - V(q_j) + q_j k_j + l_j^2/2), \quad \text{(A9)}$$

which is essentially the sum in (A8). We can neglect the term $l_j^2/2$, from the Lagrangian; this is legitimate as the action occurs as $\mathrm{e}^{2\pi i N S}$, and since N is an even integer these terms do not matter. Note that the trace formula at $T = 1$ is easily obtained from the above and has been used in the main body of this paper.

The integral is evaluated by standard methods.

$$\int_{-\infty}^{\infty} \mathrm{d}^T y \exp(i\pi N y^t A_\beta y) = \frac{\mathrm{e}^{i\pi T/4}}{N^{T/2}} \frac{\mathrm{e}^{-i\pi \nu_\beta/2}}{\sqrt{|\det(A_\beta)|}}. \quad \text{(A10)}$$

Here y denotes the T component vector $\{y_j\}$ and y^t is its transpose. The real symmetric matrix A_β is given by

$$A_\beta = \begin{pmatrix} 2 - V''(q_1) & -1 & 0 & .. & .. & -1 \\ -1 & 2 - V''(q_2) & -1 & 0 & .. & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -1 & 0 & .. & .. & -1 & 2 - V''(q_T) \end{pmatrix}. \quad \text{(A11)}$$

ν_β is a "Maslov like" index, and is the number of negative eigenvalues of the matrix A_β.

The nontrivial problem of evaluating the determinant is already solved and the results are well known [14]. The determinant can be related to the stability of the periodic orbit. If the residue of the periodic orbit, which is simply related to the trace of the stability matrix, is R_β, then $R_\beta = -\frac{1}{4}\det(A_\beta)$. or if we denote the stability matrix eigenvalues by λ_+^β and λ_-^β, $\det(A_\beta) = \lambda_+^\beta + \lambda_-^\beta - 2$.

Thus finally we can write the asymptotic periodic orbit sum, or the trace formula, for the toral maps as

$$\mathrm{Tr}(U^T) \sim \sum_{\beta} \frac{\exp\left(2\pi i N S_\beta - i\pi\nu_\beta/2\right)}{\sqrt{|\lambda_+^{\beta T} + \lambda_-^{\beta T} - 2|}}. \tag{A12}$$

The sum is over periodic orbits whose periods are integer fractions of T are labeled by β. This is of the canonical Gutzwiller-Tabor form having in the exponent the *action* of the period T orbits, and the prefactor explicitly depending on the *stability* of these orbits. The piecewise linear maps such as the baker's map, or the sawtooth map are such that the Maslov like phases are zero and the only type of periodic orbits are those of the direct hyperbolic kind. Now however we can study the vast class of *mixed* systems, such as the well known standard map. It is then of interest to ask questions such as how long the semiclassical approximations are valid in such systems which one knows are generic. The quantum mechanics on the torus is exact and thus we expect that the derivation above will facilitate the investigation of the periodic orbit sum for mixed systems. In this note we have used the above formula only in the special case of $T = 1$.

References

[1] M C Gutzwiller, *Chaos in classical and quantum mechanics* (Springer, New York, 1990)
[2] M A Sepulveda, S Tomsovic and E J Heller, *Phys. Rev. Lett.* **69**, 402 (1992)
[3] N L Balazs and A Voros, *Ann. Phys. (N.Y.)* **190**, 1, (1989)
[4] P W O' Connor, S Tomsovic and E J Heller, *Physica* **D55**, 340 (1992)
A M O De Almeida and M Saraceno, *Ann. Phys. (N.Y.)* **210**, 1 (1991)
M Saraceno, *Ann. Phys. (N.Y.)* **199**, 37 (1990)
[5] M Saraceno and A Voros, *Physica* **D79**, 206 (1994)
[6] A Lakshminarayan, *Ann. Phys. (N.Y.)* **239**, 272 (1995)
[7] F M Izrailev, *Phys. Rep.* **196**, 299 (1990)
[8] G N Watson, *A treatise on the theory of Bessel functions* (Cambridge University Press, London, 1966)
[9] J H Hannay and M V Berry, *Physica* **D1**, 267 (1980)
[10] A Lakshminarayan, *Phys. Lett.* **A192**, 345 (1994)
A Lakshminarayan and N L Balazs, *Chaos, solitons, fractals* (Osaka) **5**, 1169 (1995)
[11] J Ford, G Mantica and G H Ristow, *Physica* **D50**, 493 (1991)
[12] J P Keating, *Nonlinearity* **4**, 335 (1991)
[13] M V Berry and J Goldberg, *Nonlinearity* **1**, 1 (1988)
[14] J M Greene, *J. Math. Phys.* **20**, 1183 (1979)
[15] P Gaspard and D Alonso, *Phys. Rev.* **A47**, R3468 (1993)
[16] G Junker and H Leschke, *Physica* **D56**, 135 (1992)
[17] M Tabor, *Physica* **D6**, 195 (1983)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 2,
February 1997
pp. 537–553

The phase of the Riemann zeta function

AVINASH KHARE
Institute of Physics, Sachivalaya Marg, Bhubaneswar 751 005, India

Abstract. We, offer an alternative interpretation of the Riemann zeta function $\zeta(s)$ as a scattering amplitude and its nontrivial zeros as the resonances in the scattering amplitude. We also look at several different facets of the phase of the ζ function. For example, we show that the smooth part of the ζ function along the line of the zeros is related to the quantum density of states of an inverted oscillator. On the other hand, for $\Re s > 1/2$, we show that the memory of the zeros fades only gradually through a Lorentzian smoothing of the delta functions. The corresponding trace formula for $\Re s \gg 1$ is shown to be of the same form as generated by a one-dimensional harmonic oscillator in one direction along with an inverted oscillator in the transverse direction. Quite remarkably for this simple model, the Gutzwiller trace formula can be obtained analytically and is found to agree with the quantum result.

Keywords. Riemann zeta function; Gutzwiller trace formula.

PACS Nos 03.65; 05.30

1. Introduction

The Riemann zeta function $\zeta(s)$ of the complex variable $s = \sigma + it$ is defined for $\sigma > 1$ by [1]

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p\varepsilon \text{primes}} \frac{1}{(1-p^{-s})} \tag{1}$$

and for $\sigma < 1$ by analytic continuation. As is well known, $\zeta(s)$ has so called trivial zeros at $s = -2n$ with $n = 1, 2, 3 \ldots$. All other zeros are at complex values of s. According to Riemann's celebrated hypothesis made in 1856, the nontrivial zeros of ζ all lie symmetrically on the line $\sigma = 1/2$, i.e. $\zeta(\frac{1}{2} \pm it_n) = 0$. Riemann's hypothesis is supported by numerical tests up to very large values of t_n but mathematicians are still unable to prove or disprove it rigorously. It is also well known that there are an infinite number of zeros on the half-line $\sigma = 1/2$ [2]. These so-called nontrivial zeros are of great interest to mathematicians from a number-theoretic point of view and to physicists interested in quantum chaos and periodic orbit theory [3]. In particular, the so-called nontrivial zeros of $\zeta(s)$ exhibit an intrinsically random distribution of the GUE type. Further, assuming that the famous Riemann hypothesis is correct, the density of zeros can be shown to obey a sum rule which is analogous to the famous Gutzwiller formula for the level density. All this is taken to suggest that there may be a classical chaotic dynamical system without time-reversal invariance and t_n are the eigenvalues of the quantum Hamiltonian which is obtained by quantizing this found.

We have offered an alternative interpretation of the ζ function as a scattering amplitude and its nontrivial zeros as resonances in the scattering amplitude [7,8]. As a bonus, this interpretation directly leads us to an approximate rule for the location of the zeros with an error which is at most 3 per cent.

We have also looked in detail at several different facets of the phase of the Riemann ζ function. For example, we show that the smooth part of the ζ function along the line of the zeros (i.e. $\sigma = 1/2$) is related to the quantum density of states of an inverted oscillator. On the other hand, for $\sigma > 1/2$, we show that the memory of the zeros fades only gradually through a Lorentzian smoothing of the delta functions. The corresponding trace formula for $\sigma \gg 1$ is shown to be of the same form as generated by a one-dimensional harmonic oscillator in one direction along with an inverted oscillator in the transverse direction. Quite remarkably, for this simple model, the Gutzwiller trace formula can be obtained analytically and is found to agree with the quantum result. As far as I am aware this is the first instance of a scattering system, for which the Gutzwiller formula is exact. Before I come to these issues, I briefly review some properties of the ζ function.

2. Some Properties of the Riemann Zeta function

The Riemann zeta function as defined in Eq. (1) can also be written as

$$\zeta(s) = \sum_{n=0}^{\infty} \exp(-\beta E_n) \tag{2}$$

where $\beta = \frac{1}{kT}$, $s = \xi\beta$ and $E_n = \xi \ln(n+1)$, with the constant ξ setting the energy scale. In this form it is clearly the canonical partition function of a quantum system with an energy spectrum E_n as given above. Without any loss of generality, we set $\xi = 1$ throughout this article.

Actually, ζ can also be looked upon as a grand partition function; since any integer n is a unique product of primes $p_i (\ln(n)) = \sum_i \alpha_i \ln(p_i)$ where α_i are a set of positive integers or zero). Hence the zeta function may also be written as

$$\zeta(s) = \prod_i \frac{1}{[1 - \exp(-\beta \ln(p_i))]}. \tag{3}$$

Written in this form, it can clearly be regarded as a bosonic grand partition function with chemical potential $\mu = 0$. In this article however we shall make use of the canonical partition function interpretation of the ζ.

The quantum density of states of the system as given by (2) with the energy spectrum $E_n = \ln(n+1)$ is defined as

$$\rho(E) = \sum_{n=0}^{\infty} \delta(E - E_n) = \sum_{n=0}^{\infty} \delta(E - \ln(n+1)) \tag{4}$$

Following Jennings [9], let us derive an expression for the smooth and oscillating parts of

$\rho(E)$ for a nondegenerate spectrum. Using the fact that

$$\sum_{n=0}^{\infty}\delta(x-n) = 1 + 2\sum_{m=1}^{\infty}\cos(2\pi mx) \tag{5}$$

we obtain

$$\rho(E) = \left|\frac{\mathrm{d}G}{\mathrm{d}E}\right|\left[1 + 2\sum_{m=1}^{\infty}\cos(2\pi mG(E))\right] \tag{6}$$

Here $G(E)$ satisfies the inverse relation $n = G(E)$. For the spectrum, $E_n = \ln(n+1)$, we have $G(E) = (\mathrm{e}^E - 1)$ and hence the density of states can be written as

$$\rho(E) = \mathrm{e}^E\left[1 + 2\sum_{m=1}^{\infty}\cos(2\pi m\mathrm{e}^E)\right] \tag{7}$$

which is exponentially rising! For an exponentially rising density of states like the one here, the entropy $S(E) = E$, and the limiting inverse temperature(which is the analogue of the Hagedorn temperature [10] in high energy physics) is given by $\beta_c = \partial S/\partial E = 1$. This is not surprising in view of the composite nature of the system and the fact that the number of elementary constituents (primes) tend to increase with increasing E (corresponding to large n). A similar situation is encountered in high energy physics [10] where the hadrons may be considered to be the composite of more elementary constituents whose number also increases with energy.

Let us now mention some standard results for the Riemann zeta Function. The most important is the functional equation satisfied by it [1]

$$\zeta(s) = 2^{-s}\pi^{s-1}\sin(s\pi/2)\Gamma(1-s)\zeta(1-s) \tag{8}$$

I would like to point out the analogy [11] between this functional equation and the Kramers-Wannier duality of the canonical partition function per site for the "infinite" planar square Ising model [12] as given by

$$Z(Ks) = Z(K's)/\sinh(2K's) \tag{9}$$

with

$$\sinh(2Ks)\sinh(2K's) = 1 \tag{10}$$

It is worth emphasizing that the Kramers–Wannier duality is rather special and the fixed point of the duality equation turns out to be the critical temperature. Also note that both ζ and Z are real for real s and their sets of zeros are invariant under both duality and complex conjugation. I would also like to emphasize the striking similarity between the result of Fisher [13] on the location of all the zeros of Z on the circle $|\sinh(2Ks)| = 1$, with the Riemann hypothesis that all nontrivial zeros are on the fixed line $\sigma = \frac{1}{2}$. In last two years, duality has played a very profound role in supersymmetric gauge field theories as well as string theories [14] and it would be worthwhile to enquire if some of those ideas could also be useful in the case of ζ function.

Using the fundamental functional relationship between $\zeta(s)$ and $\zeta(1-s)$ as given by Eq. (8), it is easy to show that

$$\zeta(\tfrac{1}{2}-it) = (\pi)^{-it}\,\frac{\Gamma(\frac{1}{4}+it/2)}{\Gamma(\frac{1}{14}-it/2)}\,\zeta(\tfrac{1}{2}+it), \tag{11}$$

where $\Gamma(z)$ denotes the gamma function of the argument z. We may further write

$$\zeta(\tfrac{1}{2}+it) = Z(t)\exp(-i\theta(t)), \tag{12}$$

where $Z(t)$ is real, and $\theta(t)$ is the phase angle, with the convention that $\theta(0)=\pi$. Using Eqs. (1) and (2), it follows that

$$\exp(2i\theta(t)) = \exp(-it\ln\pi)\,\frac{\Gamma(\frac{1}{4}+it/2)}{\Gamma(\frac{1}{4}-it/2)}. \tag{13}$$

The phase θ, as defined above, is smooth in the sense that it does not include the jumps by π due to the zeros of $Z(t)$. Nevertheless, the number of zeros between 0 and t on the $\sigma = 1/2$ line is counted fairly accurately by $\theta(t)$, as will become clear from the Argand diagram. Note that

$$\frac{\theta(t)}{\pi} = -\frac{t}{2\pi}\ln\pi + \frac{1}{2\pi}\Im\left[\ln\Gamma\left(\frac{1}{4}+\frac{it}{2}\right) - \ln\Gamma\left(\frac{1}{4}-\frac{it}{2}\right)\right] + 1, \tag{14}$$

which satisfies the condition that $\theta(0) = \pi$. The density of zeros is given by

$$\frac{1}{\pi}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{1}{2\pi}\left[-\ln\pi + \Re\left[\Psi\left(\frac{1}{4}+i\frac{t}{2}\right)\right]\right], \tag{15}$$

where the digamma function is defined as $\Psi(z) = \Gamma'(z)/\Gamma(z)$. From the above, the asymptotic expression for $\theta(t)$ may be obtained immediately by making asymptotic expansion of the Γ functions. We denote this by $\tilde{\theta}(t)$, and it is given by

$$\frac{1}{\pi}\tilde{\theta}(t) = \left(\frac{t}{2\pi}\right)\ln\left(\frac{t}{2\pi}\right) - \left(\frac{t}{2\pi}\right) + \frac{7}{8} + \frac{1}{48\pi t} + \cdots. \tag{16}$$

3. Argand diagram and analogy with the scattering amplitude

To bring out some characteristics of the function $\zeta(1/2+it)$, we plot its Argand diagram in figure 1(a) in the range $t = 9$ to $t = 50$. This shows a sequence of closed loops, one for every zero of the zeta function. At a zero of $\zeta(1/2+it)$, both its real and imaginary parts are zero at the same value of t, and therefore every loop converges at the origin. The intercepts on the real axis are the so-called "Gram points" where only imaginary part of $\zeta(s)$ is zero due to the phase angle $\theta(t) = n\pi$. With infrequent exceptions, there is one Gram point between two consecutive zeros of the ζ function. The first two exceptions to this rule occur for the 126th and the 134th zeros at $t = 282.455$ and 295.584 respectively [1]. In figures 1(b) and 1(c), Argand diagrams are drawn away from the 1/2-axis, for $\sigma = 0.6$ and $\sigma = 1$ respectively. These clearly show the defocussing at the origin due to the absence of the zeros in the Zeta function. Moreover, the number of intercepts along

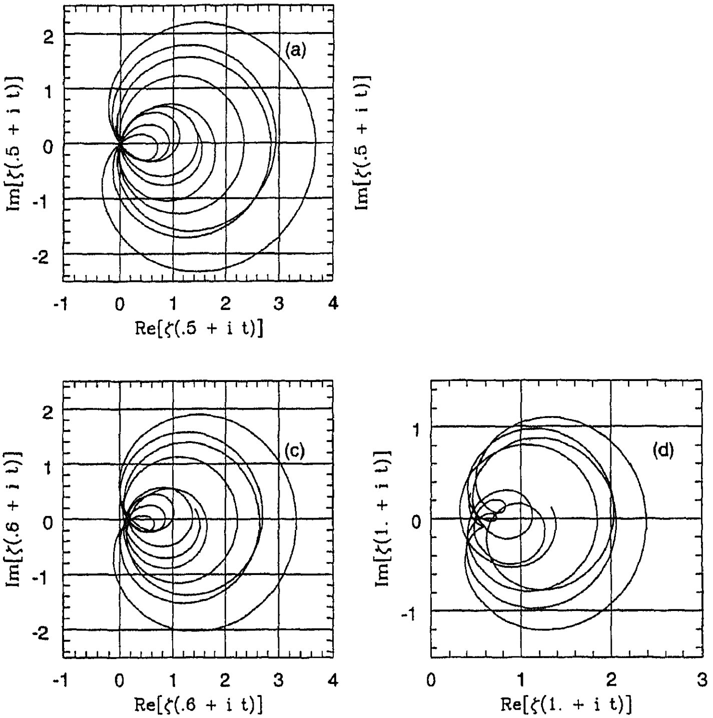


Figure 1. The Argand diagrams for $\zeta(\sigma + it)$ for increasing t and fixed σ. **(a)** $\zeta(1/2 + it)$ in the range $t = 9 - 50$. The lower limit for t is chosen so as not to miss the Gram point flanking the lowest zero at $t = 14.13$. **(b)** $\zeta(0.6 + it)$ for $t = 9 - 50$ to show the defocusing at the origin. **(c)** $\zeta(1 + it)$ for $t = 9 - 50$. Note the pronounced shift of the diagram away from the origin in this case.

the real-axis in the Argand diagrams now show a large increase compared to the $\sigma = 1/2$ case, whereas the intercepts on the imaginary axis are few or nonexistant. This is a reflection of the change in the behaviour of the phase $\theta(t)$ away from the $\sigma = 1/2$ line. In figure 2(a), the phase angle $\theta(t)$, as determined by eq. (14), is plotted as a function of t on the $1/2$-axis. This phase angle is a smooth function of t because the jumps by π at every zero (due to the change in the sign of the ζ-function) is not registered by it. These discontinuities are shown separately in figure 2(b). The smooth phase keeps increasing monotonically with t, since the curve in the complex plane passes through the origin at every zero .

Finally, in figures 3(a) and 3(b), the Argand diagrams of the Zeta function are drawn for a much larger range of t, from 1 to 500, on and off the $1/2$-axis. Note that the scale for $\sigma = 1$ is expanded compared to that for $\sigma = 1/2$. Borrowing from the terminology of the motion of a particle in phase space, it is as if there is an "attractor" at the origin for $\sigma = 1/2$ (figure 3(a)), which is absent from the more disorderly tracks of figure 3(b), which is drawn along the $\sigma = 1$ line. The latter figure also shows that the real part of the ζ function is always positive for $\sigma = 1$ for this entire range of t.

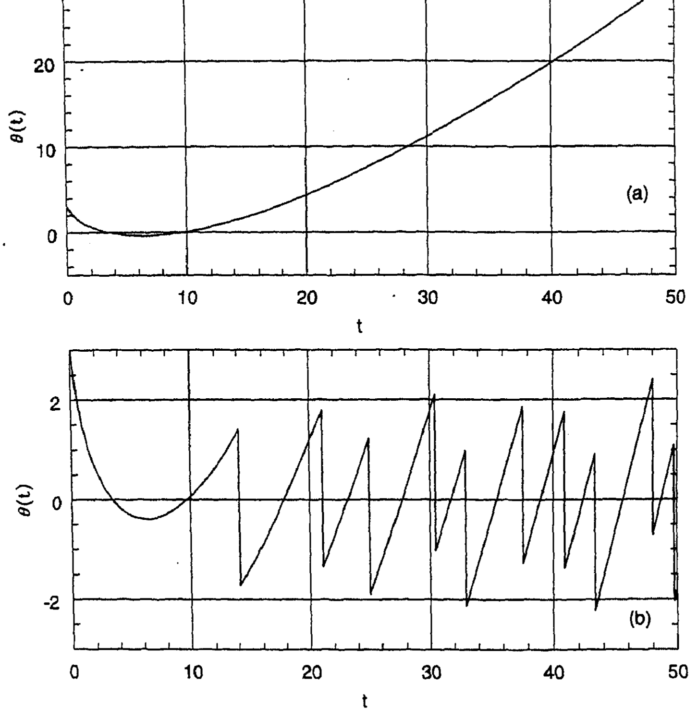


Figure 2. (**a**) The smooth phase angle of $\zeta(1/2+it)$, as defined by Eq. 14, is plotted as a function of $t = 0 - 50$. (**b**) The discontinuities of π in the phase angle $\theta(t)$ at $\sigma = 0.5$ at the position of the zeros are shown. The total phase is given by the superposition of (b) on (a).

The loop structure of the ζ function at $\sigma = 1/2$, with some near-circular shapes, is reminiscent of the Argand plots for the scattering amplitudes of different partial waves in the analysis of resonances, for example in pion-nucleon scattering [15]. Consider the partial wave amplitude $f_l(k)$, defined in terms of the partial wave phase-shift $\delta_l(k)$ and the inelasticity parameter $\eta_l(k)$,

$$f_l(k) = (\eta_l \exp(2i\delta_l) - 1)/2ik. \tag{17}$$

Here l refers to the angular momentum, and k the wave number. Note that $\Im f_l(k)$ is never negative, since the inelasticity parameter η_l is always less than one. One generally plots an Argand diagram with $2k\Im f_l(k)$ along the y–axis and $2k\,\Re f_l(k)$ along the x-axis for various values of k. For the case of no inelasticity ($\eta_l = 1$) and a single resonance, the Argand diagram is a perfect circle with unit radius, with the center on the imaginary axis at 1. By comparing this with figure 1(a) at $\sigma = 1/2$, we see that the real and the imaginary parts are interchanged in the latter, but otherwise there is a strong similarity, with many of the loops having inelasticity. This analogy is flawed, however, since $\Re\zeta(1/2+it)$ does become negative in small islands of t. Nevertheless, if these islands are ignored, then the phase shift δ_l may be identified with the phase angle $\theta + \pi/2$, with each

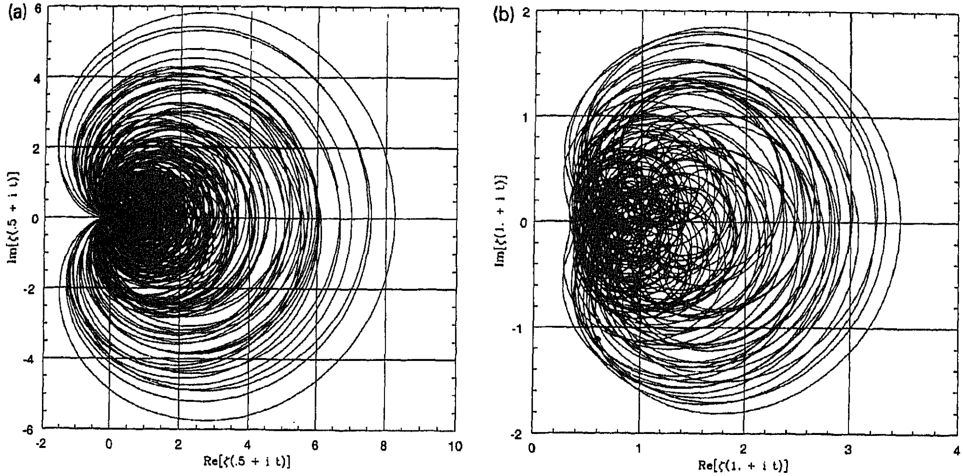


Figure 3. The Argand diagrams of the Riemann zeta function for the wider range of t−values, from $t = 9 - 500$. **(a)** $\zeta(1/2 + it)$, and **(b)** $\zeta(1 + it)$.

closed loop in figure 1(a) being regarded as in isolated resonance. In this approximation, the Gram points occur as before for $\sin\theta = 0$, while the zeros of $\zeta(1/2 + it)$ are given by the condition

$$\cos\theta = 0, \quad \theta = (m + 1/2)\pi, \quad m = 1, 2, \ldots. \tag{18}$$

This condition for the location of the zeros was also obtained by Berry [5] from the first term in his approximate formula. Eq. (18) has roots that yield the zeros on the 1/2-axis with an error of at most 3 per cent.

4. The inverted harmonic oscillator

On the $\sigma = 1/2$ line, our analogy with the scattering amplitude suggests that the phase angle $\theta(t)$ is related to a scattering phase shift. We now demonstrate that the scattering of a nonrelativistic particle by an inverted harmonic oscillator with a hard wall at the origin generates a phase shift that is closely related to $\theta(t)$. Indeed, we show that the quantum density of states for this problem is essentially the same as Eq. (14) for the density of the zeros. Consider the Schrödinger equation for $x \geq 0$,

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Phi - \tfrac{1}{2}m\omega^2 x^2\Phi = E\Phi, \tag{19}$$

and impose the boundary condition that the wave function Φ vanishes at the origin. Putting $x^2 = y$, $\Phi = y^{-\frac{1}{4}}\phi$, it becomes

$$\frac{d^2}{dy^2}\phi - \frac{l(l+1)}{y^2}\phi - \frac{kt}{y}\phi + k^2\phi = 0. \tag{20}$$

In the above equation,

$$l = -\frac{1}{4}, \quad k = \frac{m\omega}{2\hbar}, \quad t = -\frac{E}{\hbar\omega}. \tag{21}$$

This is effectively a three dimensional Schrödinger equation for a repulsive Coulomb potential in the variable y. To obtain the phase shift, we write the asymptotic solution of the above equation as

$$\phi(y) \sim \sin\left(ky - \frac{t}{2}\ln(2ky) - \frac{l\pi}{2} + \eta_l\right), \tag{22}$$

where η_l is the phase shift with respect to the distorted Coulomb wave, given by $\arg\Gamma(l+1+it/2)$. For our one-dimensional problem, only $l = -1/4$ is relevant. For this case, omitting the subscript l, the phase shift η is

$$\eta = \arg\Gamma\left(\frac{3}{4} + \frac{it}{2}\right). \tag{23}$$

$$\Gamma(\tfrac{1}{4} + iy)\,\Gamma(\tfrac{3}{4} - iy) = \frac{\pi\sqrt{2}}{\cosh\pi y + i\sinh\pi y}. \tag{24}$$

the number of quantum states $n(t)$, between 0 and t, is then given by

$$n(t) = \frac{\eta(t)}{\pi} = \frac{C}{2\pi} + \frac{1}{2\pi}\Im\left[\ln\Gamma\left(\frac{1}{4} + i\frac{t}{2}\right) - \ln\Gamma\left(\frac{1}{4} - i\frac{t}{2}\right)\right]. \tag{25}$$

In the above equation, C is a smooth function given by

$$C = \frac{\pi}{2} - \tan^{-1}(\operatorname{cosech}\pi t). \tag{26}$$

Note from above that the expression for $\eta(t)$ is not quite identical to $\theta(t)$ as defined in Eq. (14). However, their derivatives, the quantum density of states, only differ by a constant and an exponentially small term. It should also be pointed out that even if we had started with a full inverted harmonic oscillator (rather than the half-oscillator), the same conclusion would be reached, even though some nonuniqueness may arise in the choice of the boundary condition. The inverted harmonic oscillator problem has been studied by a number of authors in the past in relation to time-delay [16] and string theory [17]. No connection, however, was made to the phase of $\zeta(1/2 + it)$.

5. The phase of the Zeta function for $\sigma > \frac{1}{2}$

We define the phase of $\zeta(s)$ along the imaginary axis t for a fixed σ to be $\theta_\sigma(t)$, given by

$$\zeta(\sigma + it) = |\zeta(s)|\exp(-i\theta_\sigma(t)). \tag{27}$$

When $\sigma = 1/2$, we shall simply drop the subscript, and denote the phase by θ. Since $|\zeta(s)|$ is always positive, the discontinuous jumps due to the sign changes are included in the phase θ. This is different from the definition of the phase given in Eq. (12), where θ denoted the smooth part only. From the above definition (27), it follows that

$$\frac{\mathrm{d}\theta_\sigma}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}(\Im \ln \zeta(s)). \tag{28}$$

It is the derivative of the phase θ with respect to t that contains the information about the density of the zeros on the line $s = 1/2 + it$. We wish to examine the same derivative for $\sigma > 1/2$. To this end, it is convenient to define [1] the entire function $\xi(s)$ which has the same zeros on the complex plane as $\zeta(s)$,

$$\zeta(s) = \frac{\pi^{s/2}}{\Gamma(s/2+1)}\frac{1}{(s-1)}\xi(s). \tag{29}$$

Because $\xi(s)$ is an entire function, it may be expressed as [1]

$$\xi(s) = \xi(0)\prod_n \left(1 - \frac{s}{s_n}\right), \tag{30}$$

where s_n are the zeros of $\xi(s)$ on the complex plane. Some straight-forward algebra then yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\Im \ln \xi(s) = \sum_{n=1}^{\infty} \frac{(\sigma - 1/2)}{(\sigma - 1/2)^2 + (t - t_n)^2}. \tag{31}$$

In the above, we have assumed the Riemann hypothesis, that the only zeros of $\xi(s)$ are at $s = s_n = 1/2 + it_n$. Noting the representation of the Dirac delta function

$$\delta(t - t_n) = \lim_{\gamma \to 0}\left(\frac{\gamma}{\pi}\right)\frac{1}{(t - t_n)^2 + \gamma^2},$$

we immediately see that the derivative of the phase changes from delta function spikes to Lorentzian as we move away from the $\sigma = 1/2$ line, with a width $\gamma = (\sigma - 1/2)$. Let us denote the density of the zeros of $\xi(1/2 + it)$ for $t > 0$ by

$$\rho(t) = \sum_n \delta(t - t_n). \tag{32}$$

Then the Lorentz smoothed density is expressed as

$$\rho_\sigma(t) = \frac{\gamma}{\pi}\sum_n \frac{1}{(t - t_n)^2 + \gamma^2}. \tag{33}$$

The complete expression for the phase derivative $\mathrm{d}\theta_\sigma/\mathrm{d}t$ can now be obtained by using the asymptotic Stirling's formula for the large arguments of the Γ function. The final expression, after some algebra, is given by

$$-\frac{1}{\pi}\frac{\mathrm{d}}{\mathrm{d}t}(\Im \ln \zeta(s)) = \frac{1}{\pi}\frac{\mathrm{d}\theta_\sigma}{\mathrm{d}t} = -[\rho_\sigma(t) - \tilde{\rho}_\sigma(t)] = -\delta\rho_\sigma(t), \tag{34}$$

where

$$\tilde{\rho}_\sigma(t) = \frac{1}{2\pi}\ln\frac{([\sigma - 1/2]^2 + t^2)^{1/2}}{2\pi}, \tag{35}$$

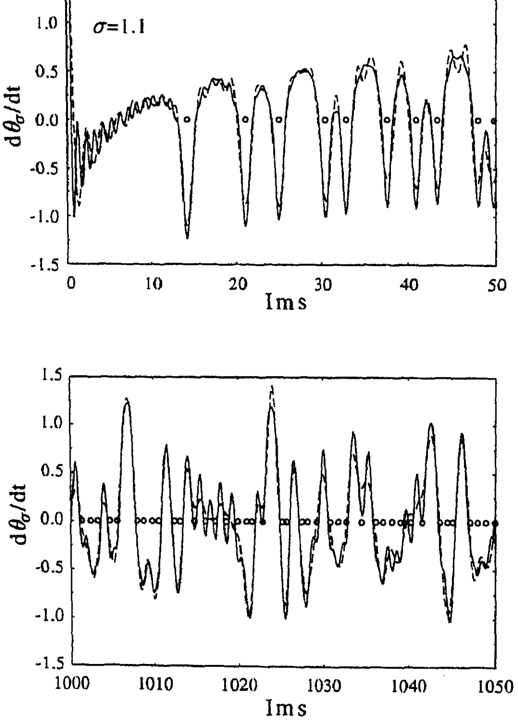


Figure 4. The derivative of the phase θ_σ of the Lorentz-smoothed Riemann zeta function, as in Eq. (34), is obtained from the trace formula, Eq. (38). It is plotted as a function of t for $\sigma = 1.1$ in the range $0 \leq t \leq 50$ (upper panel) and $1000 \leq t \leq 1050$ (lower panel). The dashed line is calculated by using the first ten primes and truncating the sum over the repetitions k at $k_{max} = 10$. The solid line includes the first 100 prime numbers, and $k_{max} = 100$. The exact position of the Riemann zeros on the $\sigma = 1/2$ axis are shown by open circles.

and ρ_σ is given by Eq. (33). In the above equation, we have neglected terms of order t^{-2} in $\widetilde{\rho}_\sigma(t)$. For $\sigma = 1/2$, Eq. (34) reduces to

$$-\frac{1}{\pi}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \delta\rho(t) = \sum_n \delta(t - t_n) - \frac{1}{2\pi}\ \ln\left(\frac{t}{2\pi}\right). \tag{36}$$

This sharply discontinuous function, according to Eqs. (33) and (34), gets Lorentz-smoothed by a width $(\sigma - 1/2)$ as the derivative of the phase is computed along the imaginary axis at $\sigma > 1/2$. Nevertheless, the position of the Riemann zeros are still remembered, with the memory gradually fading with increasing σ, as shown in figure 4.

We next consider the Euler product form for $\zeta(s)$ with a view to construct a dynamical model for large σ. This is given by [1]

$$\zeta(\sigma + it) = \prod_p \left(1 - \frac{1}{p^{\sigma+it}}\right)^{-1}, \tag{37}$$

where the product is over all the primes p. For $\sigma > 1$, the above expression is convergent. Following [4], one may then obtain the convergent trace formula

$$\delta\rho_\sigma(t) = -\frac{1}{\pi}\sum_k\sum_p \frac{(\ln p)}{(p)^{k\sigma}}\cos(kt\ln p). \tag{38}$$

For $\sigma \gg 1$, only the $p=2$ term dominates in the above trace-formula, and the contribution of the higher harmonics ($k > 1$) may be neglected. Then Eq. (38) may be written as [5]

$$\delta\rho_\sigma \simeq -\frac{1}{\omega_1}\frac{\cos(2\pi t/\omega_1)}{\sinh(\pi\omega_\sigma/\omega_1)}, \tag{39}$$

where

$$\omega_1 = \frac{2\pi}{\ln 2}, \quad \omega_\sigma = 2\sigma. \tag{40}$$

We shall now present a dynamical toy model which will generate a semiclassical trace formula, which for its lowest harmonic $k = 1$ is the same as Eq. (39). Moreover, in this situation, since only one term survives in the Euler product formula (37), the Argand diagram for $\zeta(s)$ is a circle. The analogy with the Argand diagram for the scattering amplitude is now vivid, with one *elastic* resonance only.

6. The density of states for a Toy-model

6.1 *The Gutzwiller trace formula*

In this section, we first derive semiclassically the trace formula for a particle executing unstable simple harmonic motion in one direction, perched on the edge of an inverted harmonic oscillator in a transverse direction.We also perform the quantum-mechanical calculation to test the semiclassical formula. Consider a particle described by the Hamiltonian ($m = 1$)

$$H = \tfrac{1}{2}(p_x^2 + \omega_1^2x^2 + p_y^2 - \omega_2^2y^2). \tag{41}$$

In the present problem the situation is rather simple, since there is only one primitive orbit along x. The trace formula is of the form

$$\delta\rho_{cl} = \frac{T_1}{\hbar\pi}\sum_{k=1}^{\infty}\frac{1}{\sqrt{|\det(\widetilde{M}_1^k - I)|}}\cos\left[k\left(\frac{S_1(E)}{\hbar} - \sigma_1\frac{\pi}{2}\right)\right]. \tag{42}$$

Here $T_1 = 2\pi/\omega_1$ is the period of the isolated orbit along x, and $S_1(E) = 2\pi E/\omega_1$ is the corresponding action. $\widetilde{M}_1^k$ is the 2×2-monodromy matrix, and σ_1 the Maslov index. The monodromy calculation is straight-forward, and confirms that the orbit is unstable (hyperbolic).

To establish the notation for this, and the subsequent Maslov index calculation we follow Creagh *et al* [18]. We start with a periodic orbit with coordinates given by

$\mathbf{q} = (x,y)$, and momenta $\mathbf{p} = (p_x, p_y)$. The stability of the orbit is determined by the propagation of small perturbations $\delta\mathbf{q}(t)$, $\delta\mathbf{p}(t)$ away from the periodic solution. The time evolution of these perturbations using the linearized equations of motion may be written in terms of a 4×4 matrix (called the matrizant) $X(t)$:

$$\begin{pmatrix} \delta\mathbf{q}(t) \\ \delta\mathbf{p}(t) \end{pmatrix} = X(t) \begin{pmatrix} \delta\mathbf{q}(0) \\ \delta\mathbf{p}(0) \end{pmatrix} \tag{43}$$

with the initial condition $X(0) = \mathbf{1}$. The value of the matrizant after one period (along the x-axis T_1) is the full monodromy matrix M_1:

$$M_1 = X(T_1). \tag{44}$$

Two of the eigenvalues of M_1 are unity, and the 4×4 matrix M may be written in uncoupled blocks of a 2×2 monodromy matrix $\widetilde{M}_1$ and a 2×2 unit matrix. It is this reduced matrix $\widetilde{M}_1^k$ that is given above in Eq. (42), the superscript k denoting the matrix for k-cycles. The monodromy matrix after one complete period is given by (we drop the superscript $k = 1$)

$$\widetilde{M}_1(T_1) = \begin{pmatrix} \cosh\,(2\pi\omega_2/\omega_1) & 1/\omega_2 \sinh(2\pi\omega_2/\omega_1) \\ \omega_2 \sinh(2\pi\omega_2/\omega_1) & \cosh(2\pi\omega_2/\omega_1) \end{pmatrix}. \tag{45}$$

The eigenvalues are obtained from $\det|\widetilde{M}_1 - \lambda I| = 0$ and are found to be $\exp(\pm 2\pi\omega_2/\omega_1)$. Therefore

$$\sqrt{|\det(\widetilde{M}_1 - 1)|} = 2\,\sinh\left(\frac{\pi\omega_2}{\omega_1}\right). \tag{46}$$

We next calculate the Maslov index σ_1 occuring in Eq. (42). For an unstable (hyperbolic) orbit, the Maslov index for the k^{th} harmonic is simply $k\sigma_1$. For our case, σ_1 can be calculated analytically and one can show that the Maslov index is simply $2k$.

The semiclassical trace formula given by Eq. (42) now reduces to

$$\delta\rho_{cl}(E) = \frac{1}{\hbar\omega_1} \sum_{k=1}^{\infty} (-1)^k \frac{\cos(2\pi k E/\hbar\omega_1)}{\sinh(\pi k \omega_2/\omega_1)}. \tag{47}$$

For $\omega_2/\omega_1 > 1$, the higher harmonics are severely damped, and the above equation is of the same form as given by Eq. (39). Note from Eq. (38), however, that there is an overall negative sign in the trace formula for the zeros of the Riemann Zeta function, indicating that the Maslov index for every periodic orbit, irrespective of the harmonic, is π. Our dynamical formula (47), on the other hand, has alternating sign with the harmonic number k, since our Maslov index is $k\pi$. This anomaly in the Maslov index has been noted in the literature [19]. Berry has speculated that it may not be possible to overcome this difficulty with a Hamiltonian that is separable into kinetic and potential terms.

Our next task is to perform a quantum-mechanical calculation of $\delta\rho(E)$ to check the validity of the semiclassical result.

6.2 *The quantum density of states*

We seek to obtain the quantum density of states for the two-dimensional system given by Eq. (41). Since there is no interaction coupling the two degrees of freedom, the desired density of states may be readily obtained by convoluting the quantum density of the harmonic oscillator in the x-direction with that of the inverted oscillator in the y-direction. The density of states of the harmonic oscillator with oscillator constant ω_1 is given by

$$\rho_1(E) = \sum_{n=0}^{\infty} \delta(E - (n + 1/2)\hbar\omega_1), \tag{48}$$

where the energy $E > 0$. The quantum density of states of an inverted harmonic oscillator has been examined carefully by Barton [12] and is given by

$$\rho_2(E) = -\frac{1}{\pi\hbar\omega_2} \Re\, \Psi(1/2 + iE/\hbar\omega_2), \tag{49}$$

where $\Psi(z) = \mathrm{d}/\mathrm{d}z \ln\Gamma(z)$ is the digamma function. In the above equation, we have omitted a constant term $2/\pi \ln L$ that arises from confining the system in a large box of length L. To compare with the semiclassical density of states given by Eq. (47), we convolute the quantum densities (48) and (49) for a given energy $E > 0$. This total energy E is shared between the two degrees of freedom of this conservative system. The convoluted density $\rho(E)$ of the system is

$$\rho(E) = \int_{-\infty}^{E} \rho_1(E - E')\rho_2(E')\mathrm{d}E'. \tag{50}$$

Note that the lower limit in the integration is $-\infty$, since a negative energy E' of the inverted oscillator may add up with the energy $(E - E')$ of the harmonic oscillator to still yield the energy $E > 0$ of the total system. Substituting from Eqs. (48) and (49) above, we get

$$\rho(E) = -\frac{1}{\pi\hbar\omega_2} \sum_{n=0}^{\infty} \Re\, \Psi\left(\frac{1}{2} + i\left[\frac{E}{\hbar\omega_2} - \left(n + \frac{1}{2}\right)\frac{\omega_1}{\omega_2}\right]\right). \tag{51}$$

We proceed to examine to what extent the semiclassical formula, given by Eq. (47), agrees with the quantum result, Eq. (51).

6.3 *Comparison of the semiclassical and quantum formulae*

The quantum formula given above is for the full density of states, comprising both the smooth and the oscillating parts, whereas the semiclassical result (47) is only for the oscillating part $\delta\rho(E)$. It is therefore necessary to subtract off the smooth part of the density of states from the quantum result before a comparison is made. The smooth part is most readily obtained numerically by performing a Strutinsky-type Gaussian smoothing [20] on the quantum density. The method consists of smoothing the quantum density of states, given by Eq. (51), by a Gaussian of width γ, and modulated by a curvature

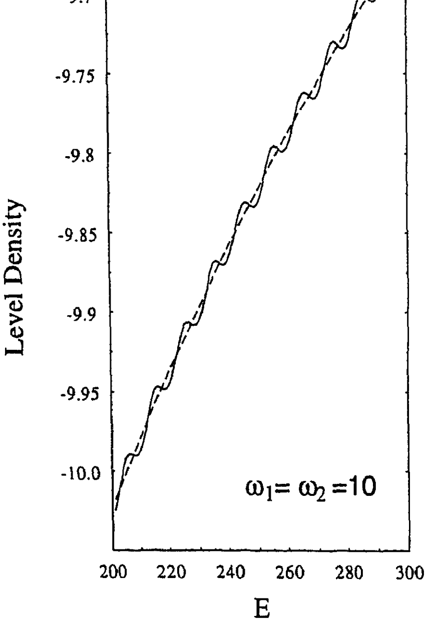


Figure 5. The quantum density of states (wavy curve), given by Eq. (51), and the Strutinsky-smoothed density, Eq. (52) (dashed line), are plotted as a function of energy E. Only the representative case $\omega_1 = \omega_2 = 10$ is shown. The smoothing width γ is chosen to be 20. The level density is negative because a large positive constant term, $2/\pi \ln L$, where $L \gg E$, has been omitted.

function $P_n[(E - E')/\gamma]$:

$$\tilde{\rho}(E) = \frac{1}{\gamma\sqrt{\pi}} \int_{-\infty}^{\infty} \rho(E - E') \exp \left(\frac{E - E'}{\gamma}\right)^2 P_n\left(\frac{E - E'}{\gamma}\right) \mathrm{d}E'. \tag{52}$$

where the smoothing width γ is larger than the spacing ω_1 of the oscillator shells. The curvature function P_n for Gaussian smoothing is given by the associated Laguerre polynomial of order $2n$:

$$P_n(x) = L_n^{1/2}(x^2). \tag{53}$$

The curvature function $P_n(x)$ ensures the internal consistency of the smoothing procedure if $\tilde{\rho}$ is a polynomial of order $(2n + 1)$. In our calculation, a curvature function with $2n = 2$ is sufficient to yield an accurate $\tilde{\rho}(E)$. The smoothing width γ was taken to be 20 and the oscillator spacing ω_1 was kept fixed at 10 (for $\hbar = 1$). The oscillating part of the density of states $\delta\rho(E)$ is obtained by taking $\rho(E) - \tilde{\rho}(E)$. This is compared with the semiclassical form (47) as a function of E for various values of the ratio ω_2/ω_1. The results of the calculation are displayed in figures 5 and 6. The agreement between the quantum and semiclassical densities is excellent, and the difference is almost negligible.

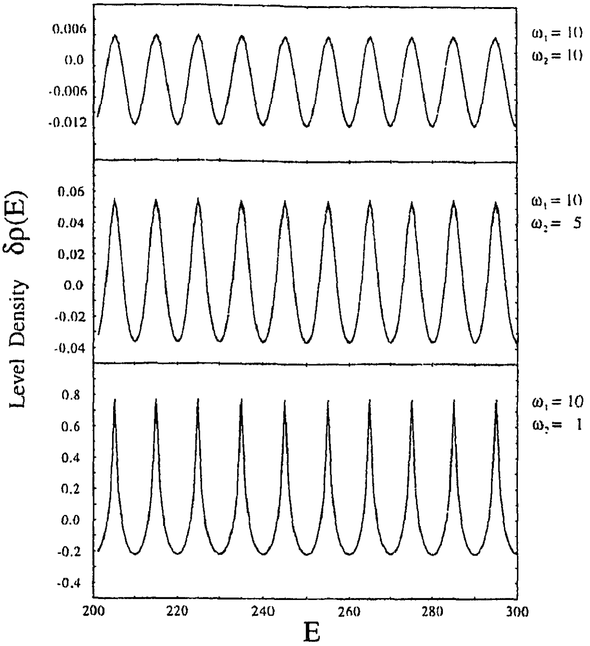


Figure 6. Comparison of the quantum and semiclassical oscillating density of states $\delta\rho(E)$, shown by the continuous and the dashed curves, for different ratios of ω_2/ω_1 given in each panel. The quantum result is obtained by subtracting the Strutinsky-smoothed density, Eq. (52), from $\rho(E)$, given by Eq. (51), with $\gamma = 20$ and $\hbar\omega_1 = 10$. The semiclassical density $\delta\rho_{cl}$ is obtained from Eq. (47) with $k_{max} = 100$.

An instructive way to analyze the semiclassical formula (47) is to construct a dynamical (or Selberg) zeta function $\zeta_S(s)$ from it, and quantize its energies by setting ζ_S to zero. Let us define

$$\zeta_S(E) = |\zeta_S(E)| \exp(-i\theta_S). \tag{54}$$

Hence we may write, following Bogomolny [21]

$$\frac{1}{\pi}\frac{\mathrm{d}\theta_S}{\mathrm{d}E} = -\frac{1}{\pi}\frac{\mathrm{d}}{\mathrm{d}E}(\Im \ln \zeta_S(E)) = \delta\rho_{cl}(E), \tag{55}$$

where $\delta\rho_{cl}(E)$ is given by Eq. (47). Note that the sign of $\delta\rho_{cl}$ in the above equation is opposite to that of the analogous Eq. (34) for the Riemann zeta function. In the latter, as was discussed in the earlier section, there is an overall minus sign in the trace formula (12) for $\delta\rho$,which may be regarded as arising from a Maslov index of π common to every periodic orbit [19]. This extra negative sign is disregarded in (55), since such a common Maslov index does not arise in the dynamical case.

From Eqs. (47) and (55), we deduce that

$$\theta_S(E) = \frac{1}{2}\sum_{k=1}^{\infty}\frac{(-)^k}{k}\frac{\sin(2\pi kE/\hbar\omega_1)}{\sinh(k\pi\omega_2/\omega_1)} = \frac{1}{2}\Im\sum_{k=1}^{\infty}\frac{(-)^k}{k}\frac{\exp(2\pi ikE/\hbar\omega_1)}{\sinh(k\pi\omega_2/\omega_1)}. \tag{56}$$

$$\frac{1}{2\sinh(x/2)} = \sum_{l=0}^{\infty} \exp -[(l+1/2)x], \tag{57}$$

and summing the series over k, we obtain

$$\theta_S(E) = -\Im \sum_{l=0}^{\infty} \ln\left(1 + \exp\left[\frac{2\pi i}{\hbar\omega_1}\left(E + i\left(l + \frac{1}{2}\right)\hbar\omega_2\right)\right]\right). \tag{58}$$

From Eq. (54), θ_S is the same as $-\Im \ln \zeta_S$, so we obtain for the Selberg zeta function the expression

$$\zeta_S(E) = \prod_{l=0}^{\infty}\left(1 + \exp\left(\frac{2\pi i}{\hbar\omega_1}(E + i(l+1/2)\hbar\omega_2)\right)\right). \tag{59}$$

The quantized energies are obtained by setting this equal to zero, giving

$$E = (n+1/2)\hbar\omega_1 - i(l+1/2)\hbar\omega_2, \tag{60}$$

where l, n take on positive integral values including zero. Note that in our case, the quantized energies E are complex. It is gratifying to see that the semiclassical Gutzwiller trace formula is so successful in describing the oscillating density of states in the continuum. To our knowledge, such an application has not been made before. It is all the more interesting because this toy-model allows us to construct a convergent dynamical zeta function. Furthermore, the model itself has some relevance, for asymptotically large σ, to the Riemann zeta function.

References

[1] H M Edwards, *Riemann's zeta function* (Academic Press, New York, 1974)
A Ivic, *The Riemann zeta function* (John Wiley, New York, 1985)
[2] G H Hardy, *Comptes Rendus* **CLVIII**, 280 (1914)
[3] M V Berry, *Proc. R. Soc. London* **A400**, 229 (1985)
M C Gutzwiller, *Chaos in classical and quantum mechanics* (Springer Verlag, New York, 1991)
Hua Wu, and D W L Sprung, *Phys. Rev.* **E48**, 2595 (1993)
[4] M V Berry, *Proc. R. Soc. London* **A400**, 229 (1985)
[5] M V Berry, in *Quantum chaos and statistical nuclear physics* (edited by T H Seligman and H Nishioka, Lecture Notes in Physics, No. 263, Springer Verlag, New York 1986) p.1
[6] M Sieber and F Steiner, *Physica* **D44**, 248 (1990); *Phys. Lett.* **A148**, 415 (1990)
A M Odlyzko, *Math. Comput.* **48**, 273 (1987)
[7] R K Bhaduri, A Khare and J Law, *Phys. Rev.* **E52**, 486 (1995)
[8] R K Bhaduri, A Khare, S M Reimann and E L Tomusiak, Bhubaneswar Preprint IP-BBSR/95-57; *Ann. Phys. (N.Y.)* (in press)
[9] B K Jennings, (1974) Unpublished
[10] R Hagedorn, *Suppl. Nuovo Cimento* **3**, 147 (1965)
[11] See for example, B L Julia, *Physica* **A203**, 425 (1994)
[12] H A Kramers and G H Wannier, *Phys. Rev.* **60**, 263 (1941)
[13] M E Fisher, *Lectures in theoretical physics* **C7**, (1965) 1, Univ. of Colorado Press, Boulder
[14] see for example, *Phys. Today*, p. 17, March 95 Issue and references therein

[15] See for example, R K Bhaduri, *Models of the nucleon* (Addison Wesley, Reading, MA, 1988) p. 27

[16] K W Ford, D L Hill, M Wakano, and J A Wheeler, *Ann. Phys. (N.Y.)* **7**, 239 (1959)
G Barton, *Ann. Phys. (N.Y.)* **166**, 322 (1986)

[17] E Brézin, V A Kazakov and Al B Zamolodchikov, *Nucl. Phys.* **B338**, 673 (1990)
D J Gross and N Miljković, *Phys. Lett.* **B238**, 217 (1990)
G Parisi, *Phys. Lett.* **B238**, 213 (1990)
P Ginsparg and J Zinn-Justin, *Phys. Lett.* **B240**, 333 (1990)

[18] S C Creagh, J M Robbins and R G Littlejohn, *Phys. Rev.* **A42**, 1907 (1990)

[19] M V Berry and J P Keating, *J. Phys.* **A23**, 4839 (1990)

[20] V M Strutinsky, *Yad. Fiz.* **3**, 614 (1966), *Nucl. Phys.* **A95**, 420 (1967)
C K Ross and R K Bhaduri, *Nucl. Phys.* **A188**, 566 (1972)

[21] E B Bogomolny, *Nonlinearity* **5**, 805 (1992)

Exotic spectra, wavefunctions and transport in incommensurate lattices

A LAHIRI, T K ROY* and A BHOWAL*
Vidyasagar Evening College, Calcutta 700 006, India
*Saha Institute of Nuclear Physics, Calcutta 700 064, India

Abstract. We present numerical results on a range of related issues for a number of incommensurate TBM's, each of which shows a metal-insulator type transition as a binding-to-hopping ratio is made to increase through some limiting value. These supplement a series of similar results on a couple of 1D lattices in a number of recent works (see below). A brief review pertaining to spectral properties and wavefunctions in incommensurate lattices is followed by results on the above TBM's relating to an interesting correlation between the gross features of wavefunctions and the energies arranged in a particular sequence termed the *lattice-ordered* sequence, and also between the lattice-ordered energies and the on-site potentials. We present a qualitative explanation of these correlations on the basis of perturbation theory. Basic results on dynamics of wavepackets in relation to spectral characteristics of incommensurate TBM's are also reviewed. Features of lattice-ordered energies and wavefunctions for the TBM's under study are used in the framework of the so-called Maryland construction, leading to a qualitative prediction of criteria for recurrent and non-recurrent wavepacket dynamics in these lattices, and these predictions are checked against numerical iterations of the relevant 'quantum maps'. Closely related to the dynamics of wavepackets are the transport properties of these lattices. Results are available to indicate that the unusual spectral characteristics of pseudorandom lattices lead to novel features in transport properties of these systems. In this context, low temperature a.c conductivity in these lattices is a good probe for the spectral characteristics and wavefunctions. However, not much is known about the a.c conductivity, excepting a set of early results pertaining to the low frequency regime, principally because of the fact that the a.c conductivity depends on global characteristics of the spectrum and the entire set of wavefunctions. We present a simple model whereby the gross structure of variation of the a.c conductivity with frequency can be obtained from a knowledge of the spectrum alone for the set of TBM's under consideration. Numerical computations show that despite its simplicity, the model leads to results in good agreement with those from the Kubo–Greenwood formula for a.c conductivity.

Keywords. Tight binding model; aperiodic lattice; localization; *m-i* transition; lattice ordered sequence; transport properties; ac conductivity; wave packet dynamics.

PACS Nos 71.50; 05.30; 71.55; 71.30; 05.45; 72.15

1. Introduction

Deterministic aperiodic lattices are interesting objects from both theoretical and applied viewpoints. Fabrication of structures of various descriptions, intermediate between amorphous and periodic ones, are now within the realm of technical possibility. The first metastable quasicrystals were grown by Schechtman *et al* [1]. Modulated 2D structures were first prepared by Weiss *et al* [2] by implanting arrays of scatterers in a semiconductor heterojunction. Techniques are now available to grow superlattices, i.e., periodically modulated structures with *large* lattice constants, in heterojunctions. Under the influence of imposed magnetic fields, electrons in these superlattices behave as if they were in a quasiperiodic environment.

Even 1D deterministic aperiodic lattice structures are now at hand, following the successful fabrication of semiconducting superstructures grown epitaxially in accordance with the rules of the Thue-Morse and Fibonacci sequences [3]. A large variety of experimental data is presently available on epitaxially grown superlattice structures on the basis of X-ray and neutron diffraction, Raman scattering, and similar other techniques (see, e.g., [3, 4] for an introduction). Concurrently, one finds extensive literature relating to theoretical studies on 1D quasirandom lattices perfomed during the last decade and a half. A large fraction of these studies pertain to simulations of electronic states resulting from tight binding models (TBM's), the initial lead in this direction having been provided by Hofstadter [5], Aubry and Andre [6], Kohmoto *et al* [7] and Ostlund *et al* [8]. These early works already established exotic features of solutions of the Schrödinger equation on a quasiperiodic lattice, namely Cantor structures of spectra, and the critical, algebraically localised, nature of wavefunctions. In this context, most studies on 1D incommensurate lattices relate to Anderson-Mott localisation in random lattices on the one hand, and to Bloch-wave solutions in periodic lattices on the other. One finds that a subtle interplay between contrary features of localised and extended states, or those of pure point and continuous spectra, indeed lends a rich variety to the physics of excitations on incommensurate lattices.

Lateral surface superlattices and 1D semiconductor heterostructures offer the welcome opportunity of *experimentally* studying the classical- to-quantum cross-over in the time-evolution of electronic states. This has caused a resurgence in theoretical and numerical investigations on quantum dynamics in situations where the underlying classical motion is irregular or chaotic. This is the area of 'dynamical quantum chaos' which initially received attention through studies on the kicked rotator [9]. Ideas of Anderson–Mott localization, when applied to results of early investigations in this area, led to the concept of 'dynamic localisation' [10], and interest in this field was about to dry up when the whole question was reopened with investigations on quantum dynamics on deterministic aperiodic structures, and with the observation that there is no direct correlation between classical chaos and quantum diffusion (see, e.g.,[11]) in these structures.

Closely related to the exotic features of excitations and wavepacket dynamics in quasirandom systems is the question of transport properties in deterministic aperiodic media. In this context, novel transport properties appear to be the order of the day, not only in relation to electronic states, but to vibrational and electromagnetic waves as well. For instance, interwoven regimes involving phonons and localised vibrational modes *(fractons)* lead to interesting effects [12, 13]. The mode of attenuation of a wave transmitted through a finite 1D chain depends sensitively on the type of eigenstates supported by the chain [14, 15]. All these observations point to interesting possibilities in fabrication of materials having new and useful properties.

Of special interest are the *a.c transport properties* of deterministic aperiodic chains. Apart from the possible future use of novelties in a.c transport in incommensurate or quasirandom systems, there is great theoretical interest in the study of a.c transport properties [16, 17]. Such a study, on the other hand, is beset with difficulties of quite enormous magnitude since the a.c transport depends on *global*, as opposed to local, features of the excitation spectrum and eigenfunctions relevant to the lattice under consideration.

We address in this paper a range of questions relating to the above context. In § 2 we point out an approximate correlation between energies and localised eigenfunctions in a number of incommensurate TBM's showing metal-insulator (m-i) type transitions. This correlation will form the basis of all subsequent discussion in the paper. Section 3 includes a short review of a few interesting results on quantum wavepacket dynamics in incommensurate systems and its correlation with the fractal properties of underlying spectra. We then present in this section a brief study on wavepacket dynamics under periodic kicking in the incommensurate TBM's considered in § 2. Section 4 addresses the question as to how far the gross features of variation of the a.c conductivity in an incommensurate lattice can be inferred from a knowledge of the spectrum alone. As we shall see, the correlation between energies and eigenfunctions presented in § 2 will enable us to infer that this can indeed be done to a certain extent, thus paving the way towards a better future understanding of this as yet largely uncharted territory. Section 5 includes summary and concluding remarks.

The work contained in the present paper is, in the main, based on a number of previous papers [18–22]. The first of these papers contained preliminary results concerning the spectrum, density of states, and localisation features in a self-similar TBM based on a universal recursive sequence encountered in Feigenbaum's theory of period doubling bifurcations [23]. The subsequent papers in the series presented results relating to a simpler self-similar TBM termed the 'Cantor lattice', and to the Aubry model. We present in this paper evidence relating to a number of other lattices admitting of m-i type transitions, to establish that the results arrived at in the earlier papers are not peculiar to the particular lattices considered therein.

2. Spectra and wavefunctions in 1D incommensurate TBM's

2.1 *Metal-insulator type transitions in 1D TBM's*

TBM's have been widely studied in the literature as useful objects mimicing real life aperiodic and incommensurate systems. In particular, despite their extreme artificiality, 1D TBM's have led to important insights based on exact results and a wealth of numerical findings. They have been fruitfully employed to explore the whole spectrum of structures ranging from periodic to random ones.

A 1D TBM with nearest neighbour hopping is typically represented by an equation of the form

$$\tau V_n \psi_n + \lambda(\psi_{n+1} + \psi_{n-1}) = E\psi_n, \tag{1}$$

where $\{V_n\}$ is some specified sequence of on-site potentials, τ a binding strength, and λ the nearest neighbour hopping strength (for introductory background on the tight binding method in solid state theory see, e.g., [24]; properties of 1D TBM's with random on-site potential sequences are presented in [25–27]). In what follows, the binding-to-hopping ratio (τ/λ) will be denoted by ν.

suitable sequences $\{V_n\}$, it admits of a set of essentially complex features like fractal or singular-continuous spectra (see Simon and Reed [28] for an explanation of the term, this reference establishes the absence of singular continuous spectra in a large class of quantum mechanical problems), self-similar and critical wavefunctions, coexisting spectra of pure point, singular-continuous and absolutely continuous varieties, and m-i type transitions.

An idea of the complexities can be had from rigorous results on the TBM with a quasiperiodic sequence of on-site potentials

$$V_n = \cos 2\pi(\alpha n + \theta), \tag{2}$$

where α is an irrational number and θ is a phase parameter. For instance, with $\nu > 2$, the spectrum has been shown to be singular-continuous when α is a Liouville number, the integrated density of states being a Devil's staircase. More generally, the singular continuous spectrum has been shown to have a Cantor set structure for a dense set of pairs (ν, θ) [29–31]. Aubry and Andre [6] established that, for irrational α, a transition from extended to localised wavefunctions occurs at the value $\nu = \nu_c = 2$ (the 'self-dual' situation) which is thus a critical parameter in a m-i type transition.

A TBM with quasiperiodic on-site potentials (2) has been termed the Aubry model in the literature. An alternative nomenclature for the incommensurate lattice, used in refs [19–22], is the quasiperiodic lattice (QPL). For the self-dual situation $\nu = 2$, it is also termed Harper's model, which represents the Schrödinger equation for an electron in a periodic potential and in a competing magnetic field, and some of whose exotic features were first pointed out by Hofstadter [5].

While the wavefunctions in the QPL with $\nu \neq 2$ are either all exponentially localised or all extended, *coexisting* states are encountered in some other models, e.g., one investigated by Soukoulis and Economou [32]. In particular, a class of 'slowly varying' potentials (see below) show mobility edges in their spectra as a common feature for a range of hopping-to-binding ratios and, at the same time, yield a m-i type transition.

2.2 *'Slowly varying' potentials and m-i type transitions*

Interesting spectral features are obtained for a number of TBM's with on-site energies possessing self-similar and hierarchical structures. Roy and Lahiri [18] investigated a TBM with a self-similar 'Feigenbaum' structure in which the sequence of on-site energies is given by

$$V_n = g^{(n)}(0), \tag{3}$$

where $g(x)$ is a well-known universal function encountered in the scaling theory of period-doubling bifurcations and $g^{(n)}$ represents the n−times iterate of g. The deterministic aperiodic lattice represented by this TBM will be termed the Feigenbaum lattice (FL).

A simpler model with a self-similar structure is the so-called Cantor lattice (CL) considered in refs [19–22], where the on-site potentials are generated by the algorithm

$$V_1 = 0; \quad V_{2n-1} = V_n/3; \quad V_{2n} = V_{2n-1} + 2/3. \tag{4a, b, c}$$

odic with periods 2^k for all integers $k > 0$, the periodicity being more larger and larger k.

CL both have a hierarchical, self-similar structure and, as indicated litatively similar results relating to spectra and wavefunctions. Other milar, hierarchical, and algorithm-generated structures have also been literature (see, e.g., refs [15], [33]). Vibrational properties of a binary or chain involve interesting phonon and fracton spectra [12–13]. While utional chains (e.g., the Fibonacci chain and the Thue-Morse chain) discussed in the literature, we shall not refer to such lattices in the

1 lattice and the Cantor lattice both show a m-i type transition, but radual and more complex than that met with in the QPL. For either at low values of the binding-to-hopping ratio (ν), the spectrum is ided states, but one finds in the spectrum sparsely distributed localised ith increasing ν the measure of localised states increases and, for above a certain limiting value (ν_c) *all* states become localized. The low which a hierarchy of mobility edges are encountered in the to lie somewhere between 5 and 6 for the FL and between 3.0 and 3.5

L and the CL have qualitatively similar structures and correspondingly tics, the latter is, in a sense, simpler than the former in that the middle ias a single scaling over the entire interval [0, 1] and is globally self-set of points $g^{(n)}(0)$ is only locally self-similar and has numerous (in ny) scale factors varying over the real line (see [18] and references details).

ence concerning the m-i type transition in the FL is to be found in [18] has been briefly indicated in [19]. The on-site potential sequences in share with the QPL the common feature of being 'slowly varying' tance, the sequence for the CL can be looked upon as a long-range periodic potential of period 2^k for sufficiently large k.

ig potential of the form

$$\text{os}\, 2\pi(\alpha n^\delta + \theta), \quad (\alpha, \theta \quad \text{as in (2)},\ \delta < 1), \tag{5}$$

y Das Sharma *et al* [34] (this being essentially the so-called 'cosine dered by Griniasty and Fishman [35] and Brenner and Fishman [36]) in a WKB type analysis, which was shown to yield correct results -i type transition observed numerically in the model: for $\nu < 2$ there bility edges separating a band of Bloch wave type extended states from onentially localised states; as ν increases towards the value $\nu_c = 2$, the ressively, till it disappears and all states become exponentially localised this paper we shall also consider a TBM based on the 'slowly varying' equence

$$\tfrac{3}{4}\cos 2\pi(\alpha n + \theta) + \tfrac{1}{4}\cos 2\pi(2\alpha n + 2\theta), \tag{6}$$

investigated by Soukoulis and Economou [32] (all the results in this paper will be for $\theta = 0$ and $\alpha = (\sqrt{5}-1)/2$ in (2), (5) and (6)). This model also shows mobility edges in the spectrum and a m-i type transition as ν is increased beyond a limiting value ν_c slightly above 1.5.

In the following, the 1D TBM's with on-site potentials given by (5), (6), will be termed the cosine lattice (CSL) and the Soukoulis-Economou lattice (SEL) respectively. Results presented in this paper will, in the main, pertain to the FL, CSL, and the SEL, while a few results on the CL and the QPL will be quoted for the sake of comparison.

Figures 1(a,b,c) indicate the m-i transitions in the FL, SEL, and the CSL respectively by showing, in each case, the variation of the inverse localisation lengths (γ) computed from the Thouless formula (see (38) in § 4 below), against energy eigenvalues computed numerically, for a number of values of the parameter ν. One observes for the FL, for instance, that for $\nu = 5.0$, there appear to be a few extended states at one edge of the spectrum while, for $\nu = 6.0$, all states appear to be localised, with ν_c apparently lying

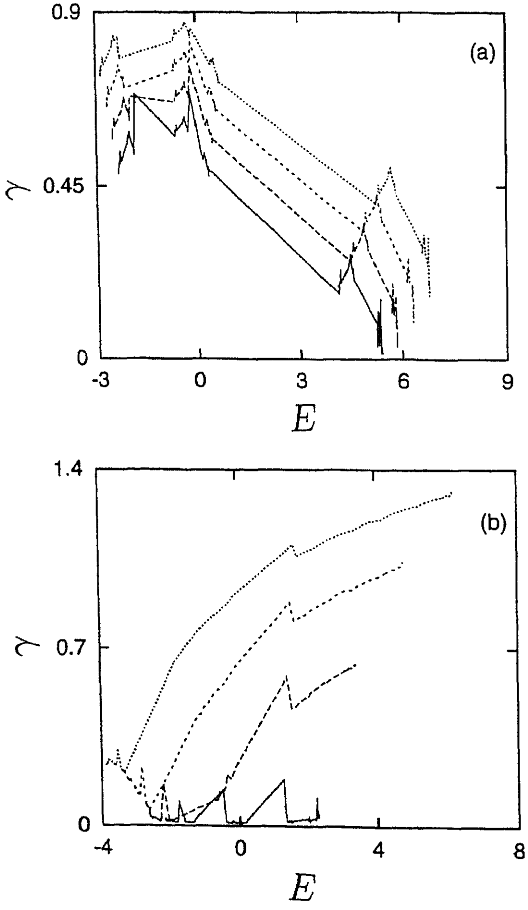


Figure 1a–b. (*Continued*)

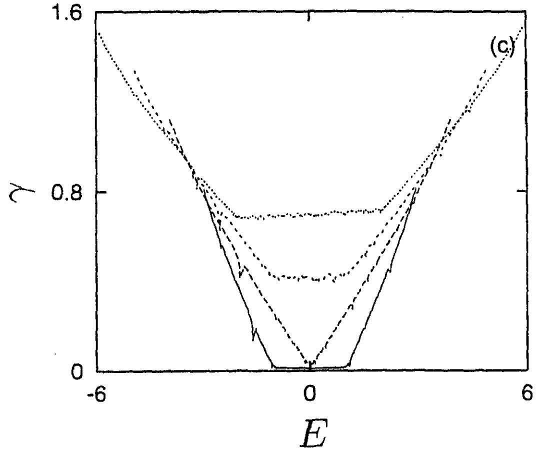


Figure 1. Inverse localization length (γ) against energy (E) for (**a**) the Feigenbaum lattice (FL) with (from bottom to top) $\nu = 5.0, 5.5, 6.0$, and 6.5 respectively, (**b**) the Soukoulis–Economou lattice (SEL) with (from bottom to top) $\nu = 1.5, 3.0, 4.5$, and 6.0 respectively, (**c**) the cosine lattice (CSL) with (from bottom to top) $\nu = 1.0, 2.0, 3.0$, and 4.0 respectively. In each of these lattices the lowermost curve includes extended states in the spectrum, the next curve is for a value of ν close to ν_c (exactly ν_c for the CSL), while the other two correspond to the localised regime. Here, as in all subsequent figures, the lattice size is $N = 1023$ and we choose $\lambda = 1.0$ (see (1)). The energies have been obtained by direct diagonalization while the inverse localization lengths have been computed from the Thouless formula.

close to 5.5. For the SEL one observes narrow bands of extended states at $\nu = 1.5$, which disappear rapidly as ν is made to increase further so that with $\nu = 3.0$, for instance, there appear to be no extended states in the spectrum. For the CSL, on the other hand, the transition is sharp and occurs at $\nu_c = 2.0$.

2.3 *Peaks of localised wavefunctions: Lattice-ordered energy sequence*

Of especial interest in this paper will be a set of results, reported in refs [19–22] for some of these models, for values of ν *beyond* the limiting value ν_c, when all the states are exponentially localized. We summarize the principal features of these results as follows:

(i) for sufficiently large ν above ν_c, the spectrum of eigenvalues is remarkably similar to the set of on-site potentials $\{V_n\}$;
(ii) there exists a certain ordering of the eigenvalues, called *lattice-ordering*, for which the locations of the peaks of the corresponding wavefunctions are strongly correlated with the energies.

Thus, the density of states resembles very closely the density of distribution of the on-site energies. For instance, figures 2(a,b) present the density of states and the density of on-site potentials for the Feigenbaum lattice with $\nu = 7.0$, wherefrom the resemblance in self-similarity between the sets of eigenvalues and of the on-site potentials is quite apparent. Significantly, the value of ν used here is not very much above the limiting

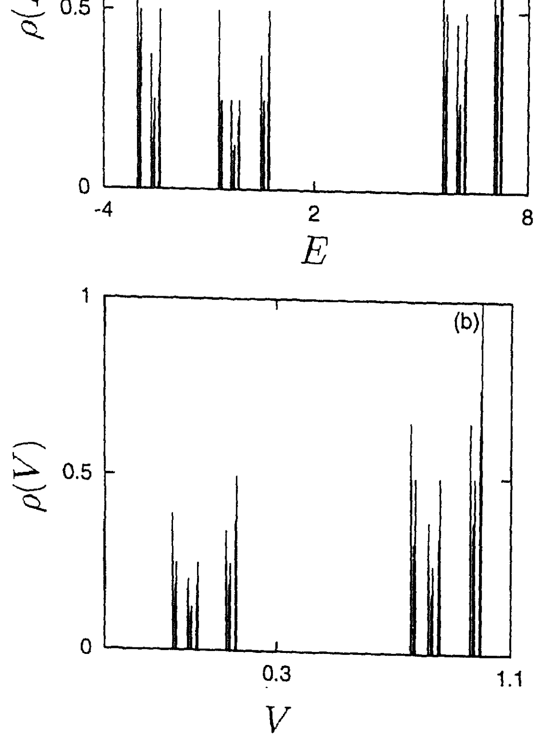


Figure 2. (a) Density of states $\rho(E)$ as a function of energy E and (b) density of on-site potentials $\rho(V)$ as a function of V for the FL with $\nu = 7.0$. Both the densities appear to be similar. Both densities are normalized to maximum value unity.

value, which means that one does not have to move far beyond the m-i transition for the above interesting features to be evident. As another instance of the similarity being referred to, arranging the set of eigenvalues in an ascending sequence $\{\tilde{e}_n\}$, and calculating $(\min_n(\tilde{e}_{n+2} - \tilde{e}_{n+1})/\min_n(\tilde{e}_{n+1} - \tilde{e}_n))$, for the CL with $\nu = 6.0$ and with even a small lattice size $N = 2^{10}$, one finds a ratio very close to 3.0, confirming that the spectrum has indeed the same scaling property as the Cantor set of on-site potentials. In particular, the entire spectrum is singular-continuous in this case as also for the FL above the m-i transition. In this respect, the situation is similar to that in case of the QPL with $\nu > 2.0$ and with α a Liouville number.

[*Remark.* A lattice like the CL or the FL actually involves an uncountable number of lattice sites and, strictly speaking, the same is true for incommensurate lattices like the

, CSL, and SEL. However, in the present paper we consider, in each case, some iciently large finite truncation of the lattice under consideration. While the orderings rred to in the following will pertain to these finite truncations, corresponding erings on the infinite lattices also exist, though they require infinite numbers of ger-valued indices for their enumeration.]

ot only do the eigenvalues resemble the on-site energies as point sets but the mblance goes even further: in particular, the two are remarkably similar when sidered as *sequences*, following a particular ordering procedure. Thus, together with sequence $\{V_n\}$ of on-site potentials describing the lattice under consideration, we sider the sequence $\{\tilde{V}_n\}$ of the same on-site potentials, but this time arranged in an nding order. Let the integer valued function $\mu(n)$ of the integer variable n give the ping between these two sequences in the sense that

$$V_n = \tilde{V}_{\mu(n)}. \tag{7a}$$

further the energy eigenvalues, arranged in the ascending sequence, be denoted by }. Then the required ordering of eigenvalues, to be called the *lattice-ordered* uence, is the sequence $\{e_n\}$ defined by

$$e_n = \tilde{e}_{\mu(n)}. \tag{7b}$$

ther words, the same shuffling that yields $\{V_n\}$ from $\{\tilde{V}_n\}$ is used to generate $\{e_n\}$ n $\{\tilde{e}_n\}$.

s already stated, the two sequences $\{e_n\}$ and $\{V_n\}$ resemble each other strongly. s, for the QPL, the variation of the lattice-ordered energies $\{e_n\}$ with n is itself found e quasiperiodic (see, e.g., [20], figure 2), exactly similar to the variation of the on-site ntials.

hat is more important, features of the wavefunctions corresponding to the energies nged in the lattice-ordered sequence are strongly correlated to the energies nselves. One of the more unusual features of incommensurate systems is that the efunctions vary extremely erratically as one moves through the energy scale. In icular, gross features of the wavefunction like the locations of the peaks or their lisation lengths have little correlation with the energies when the latter are arranged he ascending or any other arbitrarily chosen order. For instance, figure 3 is a plot of tions $N(n)$ of peaks of the wavefunctions correponding to energies $\tilde{e}_n$ arranged in the nding sequence against the sequential order n of the energies, for the QPL in the lized regime with $\nu = 4.$, which show that $N(n)$ is an extremely violently fluctuating ction of n.

correlation, however, is most strikingly revealed when one observes that *the peak of wavefunction $\psi^{(n)}$ corresponding to the energy e_n arranged in the lattice-ordered uence is precisely at the lattice site n,* with only few exceptions. In other words, oting the location of the peak of $\psi^{(n)}$ by $\mathcal{N}(n)$ one has, to a high degree of accuracy,

$$\mathcal{N}(n) = n. \tag{8}$$

ures 4(a,b) are plots of $\mathcal{N}(n)$ vs. n for the FL and the SEL, both at values of ν above respective limiting values ν_c. While the correlation expressed by (8) is found to hold arkably well in these cases, figure 5(a) is the same plot for the CSL, where one finds

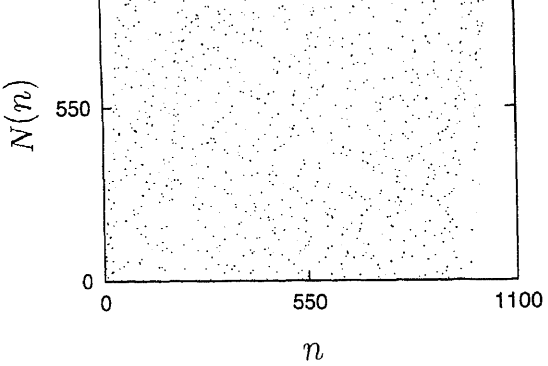


Figure 3. $N(n)$ vs. n plot for the QPL with $\nu = 4.0$. Here $N(n)$ stands for the peak of the wavefunction corresponding to energy $\tilde{e}_n$, where $\{\tilde{e}_n\}$ denotes the sequence of energy eigenvalues arranged in the ascending order. The points appear to be scattered throughout the plot, indicating an erratic variation of N with n implying that there is no correlation between the energies arranged in the ascending sequence and features of the corresponding wavefunctions. The wavefunctions have been obtained by numerical computation of eigenvectors.

that the correlation (8) holds only rather weakly. Finally, figure 5(b) is the $\mathcal{N}(n)$ vs. n plot for a *random* lattice, where the correlation is found to be *altogether absent.*

A first-hand explanation of the correlations noted above can be had from a simple perturbation theoretic argument. Thus, a TBM Hamiltonian can be written as

$$H = \tau(H_0 + \nu^{-1}H_1), \tag{9}$$

where H_0 is diagonal with diagonal elements forming the sequence $\{V_n\}$, and H_1 is given in the corresponding basis by

$$H_1 = | n+1\rangle\langle n | + | n\rangle\langle n+1 |. \tag{10}$$

When H is written in the form (9), ν^{-1} can be used as a convenient perturbation parameter, and for $\nu^{-1} = 0$ (infinitely strong binding), the eigenvalues of H are just the V_n's, with the $| n\rangle$'s being the corresponding eigenvectors. In other words, in this limit, the lattice-ordered sequence of eigenvalues is trivially the *same* as the sequence of on-site potentials, and $\psi_m^{(n)}$ is just δ_m^n, so that (8) is also trivially satisfied. For small and finite ν^{-1} one has now to see the effect of the perturbation $\nu^{-1}H_1$.

The first observation is that the perturbation theory is applicable only in the regime of localized states since the uncorrected states are all localized, and that $\nu^{-1}H_1$ cannot at all be looked upon as a perturbation unless the corrections as predicted by the perturbation series (see, e.g., [38]) are actually small. Thus, noting that the first order correction to the energies is zero, and considering the second order correction, since $\langle n | H_1 | m\rangle$ is non-zero only when $m = n \pm 1$, the validity of perturbation theory depends on how the successive uncorrected energies in the lattice-ordered sequence, i.e., the V_n's, are clustered together. It is precisely here that the incommensurate lattices we are considering in this

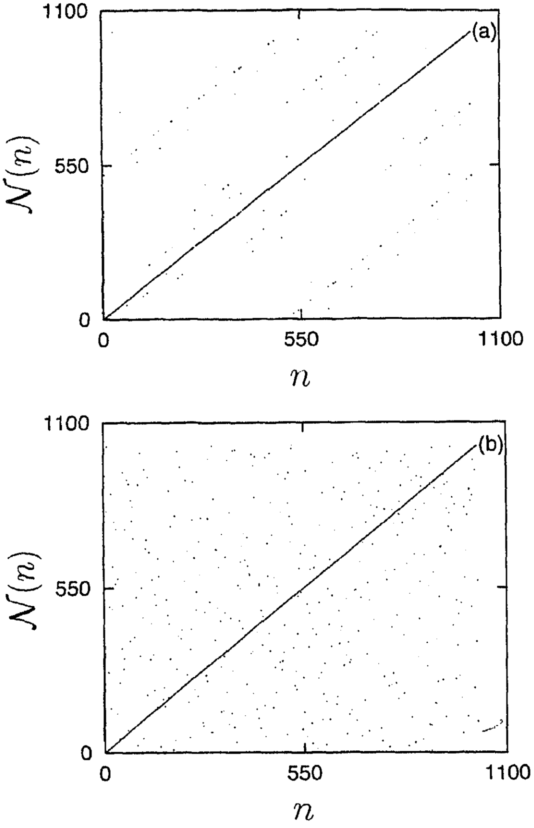


Figure 4. $\mathcal{N}(n)$ vs. n plot for (**a**) the FL with $\nu = 7.0$ and (**b**) the SEL with $\nu = 6.0$ (see § 2 (8) for explanation). Plots show a strong correlation between the energies arranged in the lattice-ordered sequence and the location of peaks of corresponding wavefunctions. Of the total number of data points in (a), 85.7% fall within the band $\mathcal{N} = n \pm 5$, while the corresponding figure for (b) is 60.2%.

paper differ radically from the random lattice. Thus, for the Cantor lattice, there is a recurrence in the V_n's only at *large* separations, and successive V_n's differ considerably from one another, so that the second order perturbation correction is always small, whereas in the case of the random lattice, there is always a considerable fraction of neighbouring V_n's differing by small margins.

This conclusion is not modified when one goes over to higher order perturbation corrections. A typical term in the perturbation series has a denominator involving differences of the form $(V_n - V_m)$ which, for the CL, is small only when $| m - n | \sim 2^k$, for k large. However, every such small denominator is offset by an immensely smaller numerator of the form $(\nu^{-1})^{2^k}$, so that the perturbation expansion is never vitiated. The situation is entirely different for the random lattice, for which there is no long range correlation among the V_n's, and the perturbation series makes no sense.

Thus, for the incommensurate lattices we are considering, the perturbation theory is applicable in the regime of localised states and so, the sequence of lattice-ordered

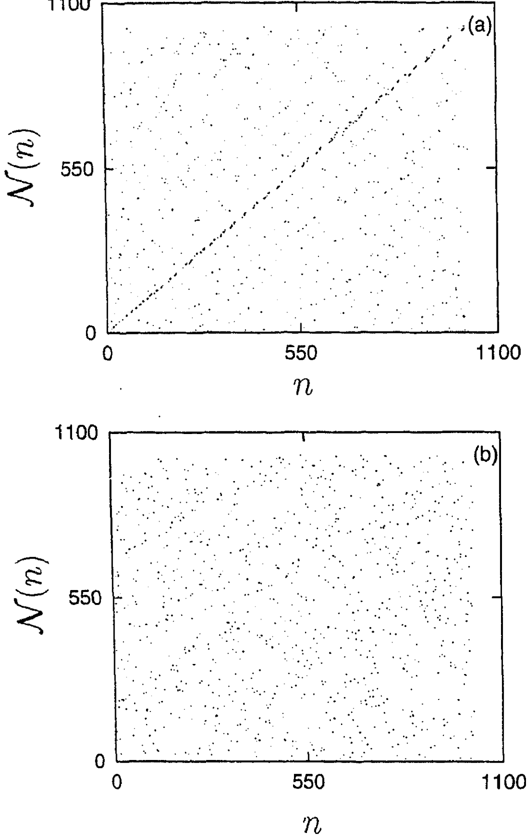


Figure 5. $\mathcal{N}(n)$ vs. n plot for (**a**) the CSL with $\nu = 6.0$, and (**b**) a random lattic with $\nu = 5.0$. In (**a**) the correlation between the energies in the lattice-ordere sequence and the location of the peaks of the corresponding wavefunctions is found t be weaker compared to correlations for the FL, the QPL and the SEL found i figure 5, only 24.8% of data points lying within the band $\mathcal{N} = n \pm 5$, while in (**b**) th correlation appears to be altogether non-existent. The random lattice was generate with a programme yielding numbers randomly in the interval (0,1).

energies, which arise from the sequence of on-site potentials continuously and withou level crossings, can indeed be expected to be qualitatively similar to the latter sequenc itself.

An essentially similar argument applies to the case of the eigenfunctions. Here agai the uncorrected functions given by $\psi_m^{(n)} = \delta_m^n$ serve as good starting points for the actu eigenfunctions for the incommensurate lattices but not for the random lattice. Since th effect of the perturbation is continuous in the strength of perturbation and since th location of the peak is an integer valued functional on the wavefunction, it is reall exceptionally stable against the perturbation, and this is what we find in figures 4 and

We close this section with the remark that the differences in degree to which the abo correlations are found to hold for the different incommensurate lattices considered in th paper (and in the previous series of papers, refs [19–22]) are related to the differences the extent to which the values of the on-site potentials V_n are clustered together. More detailed numerical results in this respect will be presented in a future publication.

These correlations will form the basis of the results and observations to follow.

3. Dynamics of wavepackets and quantum non-recurrence

3.1 *Recurrent and non-recurrent quantum dynamics*

The comparison and correspondence between classical and quantum time-evolutions has always been a keenly studied subject. While 'most' classical Hamiltonians are non-integrable, and chaotic trajectories are encountered generically in classical motions, quantum dynamics governed by the Schrödinger equation is, by contrast, *integrable* [39, 40]. However, this 'simplicity' in quantum dynamics is offset by the fact that the *state space is infinite dimensional.* As will be seen below, this gives rise to an essential complication in quantum time evolution even in a time-independent problem, as soon as one tries to look into the evolution of a wave-packet composed of stationary states.

An essential characteristic of classical chaos is that it is manifested only as an *asymptotic* feature of the motion in which a trajectory explores the phase space on a finer and finer scale in progressively larger time intervals. Quantum interference rules out a corresponding occurrence as soon as a non-zero $\hbar$ appears in the theory. This is usually interpreted as the 'absence of quantum chaos'. The presence of chaos in the classical regime and its absence - in the above sense - in the quantum regime do not contradict the quantum-classical correspondence principle because of the fact that the limits $\hbar \to 0$ and $t \to \infty$ do not commute with each other (see e.g., [40]).

However, even though an infinitely detailed exploration of the phase space is ruled out in the quantum regime, there is nothing in principle to rule out a similarly detailed exploration of the *Hilbert space*, and *this* may quite possibly lead to a complicated dynamics peculiar to the quantum regime. In any case, the phase space as such is not a well-defined or useful concept in the quantum regime and as we shall see below, wavepackets do indeed evolve in certain situations in a diffusive manner, which may be looked upon as some sort of 'chaotic' dynamics, especially so since it entails a decay of the relevant auto-correlation function. Such a diffusive quantal time evolution occurs when there are accumulation points in the underlying spectrum and involves progressively detailed and fine exploration of the spectrum with the passage of time. It is here that the infinite dimensionality of the Hilbert space plays a crucial role. Recent years have witnessed a shift towards this area of discourse in investigations on dynamic quantum chaos. Thus, while earlier studies looked principally into possible signatures of classical chaos in quantum time evolution, more recent studies are based on a recognition of the distinctness of the two regimes and of essentially independent criteria, involving distinct time scales, governing classical and quantum chaos.

The two traditional areas of investigation in 'quantum chaos' have been the 'static' and the 'dynamic' problems respectively - the former dealing with signatures of classical chaos in quantum spectra and stationary wavefunctions and the latter with the possible existence of chaos in quantal time evolution. While we have singled out the latter area in the above discussion and will, in the remainder of this section, continue to dwell mainly set of parameter values characterizing the Hamiltonian. As already mentioned, most results in these investigations pointed to what is now known as quantal suppression of classical chaos or, equivalently, 'dynamic localization'. This essentially means that asymptotically, i.e., for large times, a wavepacket evolves in an effectively quasiperiodic manner. This is the hall-mark of integrable dynamics, and is reminiscent of KAM-type time-evolution in classical phase space. More precisely, a typical wavepacket first undergoes an apparently diffusive spread reflecting the chaos in the underlying classical dynamics but, after the lapse of a characteristic time scale, there occurs a cross-over to a typically quasiperiodic regime. The problem of quantal time-evolution can be mapped to an equivalent stationary problem on what may be termed the 'energy lattice', and can then be discussed in the framework of Anderson-Mott theory of localization. If this equivalent problem yields localized states on the energy lattice then the time-evolution of the wavepacket in the original dynamical problem may be taken to be quasiperiodic. It is in this sense that one talks of 'dynamic localization' in quantum time evolution. In this case the set of stationary states contributing to the evolving wavepacket fluctuates with time within limits, as a result of which the wavepacket sort of reassembles itself, though the process of re-assembly is quasiperiodic rather than periodic. We term such a time-evolution 'recurrent', as opposed to 'non-recurrent' time-evolution where the wavepacket does not re-assemble itself even approximately at any point of time. Non-recurrent dynamics can in turn be of essentially two distinct types, namely diffusive and ballistic, the latter resembling a freely propagating wavepacket while the former resembles what is known as 'mixing' motion in classical dynamics, the autocorrelation function of the evolving wavepacket decaying more slowly than in the case of ballistic motion.

As we shall indicate in this section, these different types of time-evolution are intimately related to characteristics of the underlying energy or quasi-energy (see below) spectra, and it is in this context that dynamics in pseudorandom systems acquires a crucial significance. In particular, the possibility of singular-continuous spectra and co-existing spectral measures of distinct types makes the problem of dynamics in such systems a highly non-trivial one, and one finds that the negative results (i.e., the ones implying the absence of quantum non-reccurence) encountered in the context of the kicked rotator are by no means of a representative nature. A number of studies on 1D incommensurate lattices have already established that subtle novelties are in store in this area. Moreover, the technical possibility of actually fabricating 1D and 2D superlattices-instances of the so-called mesoscopic systems-makes the classical-quantum cross-over regime an experimentally accessible one. All this has made the problem of quantum time evolution in incommensurate systems an altogether interesting study.

3.2 *Spectral characteristics and time evolution: The quasi-energy spectrum*

The simplest of quantum time-evolutions is that of a stationary state of a time-independent system. In principle, the time evolution of an arbitrary state or a wavepacket for a system described by a time-independent Hamiltonian is a superposition of harmonically evolving terms; but whether this superposition leads to a recurrent or non-recurrent time evolution depends on the set of frequencies involved, i.e., on the energy spectrum. We confine our discussion to quantum dynamics in random and pseudorandom systems, referring, in particular to 1D TBM's for illustration.

A pure point or discrete spectrum results in a recurrent time evolution where the wavepacket pulsates regularly and its mean squared spread does not increase unboundedly with time. This is the situation in a 1D random lattice: the wavepacket does not diffuse and if the wavepacket is initially localised on a particular site, then the probability of finding it at the same site at a later time t does not go to zero as $t \rightarrow \infty$ (see e.g., [26]).

On the other hand, the presence of a singular continuous or absolutely continuous part in the spectrum complicates matters. If the initial wavepacket belongs to the subspace built up from the absolutely continuous and singular-continuous parts of the spectrum then it becomes a scattering state, i.e., ultimately escapes to infinity [29]. This is the case of nonrecurrent dynamics. It includes, in turn, a variety of different possible types of motion differing from one another in subtle details.

For instance, a periodic lattice with an absolutely continuous spectrum forming a band yields a ballistic motion, with the mean squared spread

$$\Delta^2(t) \equiv \sum_n (n - n_0)^2 \mid \psi_n(t) \mid^2, \tag{11}$$

of the wavepacket varying asymptotically as the square of the time elapsed : $\Delta^2(t) \propto t^2$ (here n_0 is the mean position of the wavepacket and ψ_n denotes the component of the wavepacket on lattice site n).

The role of singular-continuous spectra in leading to non-recurrent motions and the correlation between multifractal properties of these spectra and various characteristics of the non-recurrent motions have drawn intense attention in recent years. Guarneri [41, 42] related the dynamics of wavepacket spreading to the fractal properties of the spectrum and showed that for unbounded diffusion of the wavepacket to take place it is sufficient that the information dimension of the spectrum be larger than 0.5. Geisel, Ketzmerick and Petschel [43] made a numerical investigation of wavepacket spreading in Harper's model which is known to possess a singular-continuous spectrum and found that

$$\Delta^2 \propto t^{\chi_1}, \tag{12}$$

with χ_1 close to 1 for $\alpha =$ 'golden mean' and other strongly irrational values, in close conformity with the earlier results of Hiramoto and Abe [44]. The spectrum in Harper's model exhibits what Geisel *et al* call 'level clustering', with the probability density $p(s)$ of nearest neighbour level spacings (s) behaving as $p(s) \sim s^{-\beta}, \beta = 3/2$. Such level clustering is typical of singular-continuous spectra, as opposed to 'level repulsion' in random matrix ensembles. The exponent β is related to the Hausdroff dimension of the spectrum as $D_{\mathrm{H}} = \beta - 1$, and Geisel *et al* conjectured that

$$\chi_1 = 2D_{\mathrm{H}}, \tag{13}$$

which implies, for Harper's model, a spreading of wavepackets through what is known as 'normal' diffusion. Since the information dimension is bounded from above by the bound established by Guarneri.

While typical wavepackets in the Harper model spread through normal diffusion, other 1D pseudorandom lattices exhibit 'anomalous' diffusion with various different exponents $\chi_1(< 1)$. Thus, in the Fibonacci lattice, it varies with the binding-to-hopping ratio ν, decreasing from a value close to 1 (normal diffusion) for low ν to $\chi_1 \approx 0.25$ for $\nu \approx 10$ [45]. For the Thue–Morse chain, on the other hand, there is a preponderance of extended states in the spectrum and the spreading is 'superdiffusive' in nature, the exponent χ_1 decreasing from a value close to 2 (ballistic motion) to close to $3/2$ for large ν (Katsanos *et al* [46]).

Apart from the exponent χ characterising the spread of the wavepacket, another quantity characteristic of the nature of quantum time evolution is the large time behaviour of the temporal autocorrelation $C(t)$ of the wavefunction $\psi(t)$:

$$C(t) = (1/t) \int_0^t \mid \langle \psi(0) \mid \psi(t') \rangle \mid^2 \mathrm{d}t'. \tag{14}$$

While pure point spectra are distinguished by non-decaying autocorrelations, the situation is more complex for singular-continuous and absolutely continuous spectra. A decay of the autocorrelation faster than any power law can be ascribed to an absolutely continuous spectrum, but the converse does not appear to be necessarily true: there are absolutely continuous spectra for which the decay follows a power law. Consequently, Avron and Simon [47] distinguished between two types of absolutely continuous spectra-the so-called transient and recurrent varieties. For instance, the Aubry model with $\nu < 2$ is characterised by a decay of the autocorrelation of the form

$$C(t) \propto t^{-\chi_2}, \tag{15}$$

where the exponent χ_2 is seen to be nearly 0.84, and this slow decay has been ascribed to a recurrent absolutely continuous spectrum. For $\nu = 2$, on the other hand (Harper's model), $\chi_2 \simeq 0.14$, pointing to a singular-continuous spectrum [45].

The classical analog of Harper's model is known to be *integrable* with regular KAM-like motion and so, the diffusive spread of wavepackets in the Harper model, with its attendant decay of the autocorrelation function, is already sort of a surprise. It shows that there is really no simple tie-up between recurrence/non-recurrence of quantum time evolution and regularity/chaos in the underlying classical motion.

A more realistic model of an electron in a lateral surface superlattice, subjected to a magnetic field, leads to a classical analog that is non-integrable and is characterized by chaotic orbits [48,49]. It is thus a matter of considerable interest to look into how the onset of classical chaos affects level clustering and the characteristics of wavepacket evolution, such as the exponents χ_1 and χ_2 introduced above.

The most widely studied system in this context is the periodically kicked Harper model (KHM), which also has a classically chaotic phase space [50].

Here the Hamiltonian has a periodic time-dependence and the quantum time evolution is governed by what is known as the *quasi-energy spectrum*. Briefly, if T be the time period of kicking, then the evolution operator $\hat{U}(t + T, t)$ propagating the state through a time interval T contains all the relevant information concerning asymptotic time evolution. This is the so-called Floquet operator which is unitary and whose eigenvalues are all of the form $\exp(-i\epsilon T)$ (we set $\hbar = 1$). The real numbers ϵ constitute the quasi-energy spectrum of the problem. There are special evolving states with quasiperiodic time evolution given by

$$\psi(t) = \exp(-i\epsilon t)\phi(t), \tag{16}$$

where $\phi(t)$ is periodic with period T. These are known as quasi-stationary states and are the generalisations of stationary states relevant for a periodically driven system.

The role of quasi-energy spectra in determining the nature of time evolution of wavepackets for a periodically driven system is similar to that of the energy spectra in case of time independent systems. Thus, pure point quasi-energy spectra lead to recurrent dynamics manifested through both the mean squared spread of the wavepacket and the fluctuations of the autocorrelation (see [51, 52]), while the time evolution is non-recurrent for absolutely continuous and singular-continuous quasi-energy spectra [53].

Studies on the KHM show that a rich variety of time evolution results for different possible choices of the parameters characterizing the model and that unbounded diffusion in the classical regime is quite different compared to that in the quantum regime, the latter arising due to the hierarchical level clustering in the spectrum. Artuso *et al* [54] have confirmed that the relation (13) between the exponent χ_{I} and the fractal dimension D_{H} remains valid for the KHM. The 'phase diagram' in the parameter space demarcating different types of wavepacket evolution has been studied in Artuso *et al* [11]). One feels that a definitive understanding of the KHM has not yet been arrived at in the literature.

3.3 *Periodic kicking in TBM's: Wavepacket dynamics in energy space*

Thus, an understanding of the nature of wavepacket dynamics for a periodically driven system requires a knowledge of its quasienergy spectrum. This can be had by making use of the so-called Maryland construction [55, 56], whereby a problem with periodic delta function kicking is mapped to an equivalent stationary problem on an 'energy lattice'(see [20]). We illustrate this in the context of a tight binding Hamiltonian with periodic kicking of the form

$$\hat{H}(t) = \hat{H}_0 + \hat{V} \sum_{s=-\infty}^{\infty} \delta(t - sT), \tag{17}$$

where $\hat{H}_0$ is a time independent TBM Hamiltonian with energy eigenvalues E_n and corresponding eigenvectors $\mid n\rangle$:

$$\hat{H}_0 = \sum E_n \mid n\rangle\langle n \mid . \tag{18}$$

In this notation, the index n no longer represents site in the actual 'real space' lattice. A typical quasistationary state with quasi-energy ω (in units of $\hbar$) can be written as

$$\mid \psi_\omega; t\rangle = \mathrm{e}^{-i\omega t} \mid u_\omega; t\rangle, \tag{19a}$$

where

$$\mid u_\omega; t + T\rangle = \mid u_\omega; t\rangle. \tag{19b}$$

and the Hermitian operators $\tilde{V}$, $\tilde{H}_0$ through the relations

$$\exp(-i\hat{V}) = (I + i\tilde{V})(I - i\tilde{V})^{-1}, \tag{21a}$$

$$\exp(-i\hat{H}_0) = (I + i\tilde{H}_0)(I - i\tilde{H}_0)^{-1}. \tag{21b}$$

Then, further defining $| \bar{u}\rangle$ through

$$| u^{(-)}\rangle \equiv (I + i\tilde{H}_0) \mid \bar{u}\rangle, \tag{22}$$

and expanding $| \bar{u}\rangle$ on the energy lattice

$$| \bar{u}\rangle = \sum_n u_n \mid n\rangle, \tag{23}$$

one arrives at an equivalent description

$$(c_n + \langle n \mid \tilde{V} \mid n\rangle)u_n = -\sum_{m\neq n}\langle n \mid \tilde{V} \mid m\rangle u_m\,, \tag{24a}$$

where

$$c_n = \tan[T(\omega - E_n)/2]. \tag{24b}$$

If $\hat{V}$ is chosen such that $\tilde{V}$ has matrix elements $\tilde{V}_{mn}$ only for $m = n \pm 1$ and, further, all matrix elements between neighbouring states are chosen to be same (say, v) we get

$$c_n u_n + v(u_{n+1} + u_{n-1}) = 0, \tag{25}$$

which is an equivalent stationary TBM with on-site potentials c_n given by (24b) (termed the 'tangent potential'). A quasienergy ω is determined implicitly by the requirement that this equivalent TBM possesses a zero eigenvalue. The u_n's determined by the corresponding eigenvector of the equivalent TBM are essentially the components of the quasi-stationary states along the energy eigenstates $| n\rangle$ of the original TBM.

For large T the tangent potential is unbounded and quasi-stationary states are localised in the energy lattice. For smaller values of T on the other hand, the spectrum and eigenfunctions of the equivalent 'tangent' lattice are determined by the parameter ν since the effective binding-to-hopping ratio is $\nu = 1/v$.

The above construction was employed in [20] to the case of an electron in a CL or a QPL derived periodically by δ-function impulses in order to arrive at a useful scheme for determining the nature of time evolution of wavepackets. The TBM one starts with can be written as

$$\hat{H}_0 = \tau\sum_n V_n|u_n\rangle\langle u_n| + \lambda\sum_n(|u_n\rangle\langle u_{n+1}| + |u_n\rangle\langle u_{n-1}|), \tag{26}$$

where we assume that boundary effects can be ignored and where the $|u_n\rangle$'s are basis vectors defining the lattice (to be designated the 'real space lattice' in the present denoting the component of $| u_n\rangle$ at the lattice site m.

this TBM Hamiltonian we add a periodic kicking term so as to arrive at

$$\hat{H} = \hat{H}_0 + \hat{v} \sum_n \delta(t - sT), \tag{28}$$

e

$$\hat{v} = \sum_n v(|u_n\rangle\langle u_{n+1}| + |u_n\rangle\langle u_{n-1}|), \tag{29}$$

ioting the strength of kicking.

t now $|\omega_n\rangle$ denote the eigenvector of $\hat{H}_0$ belonging to the eigenvalues $= 1, 2, \ldots$), $\{e_n\}$ being the *lattice-ordered* sequence of eigenvectors. For the lattices of the slowly varying type introduced in § 2, we already know from the rvations in § 2 that for large enough ν (above the relevant limiting value ν_c) $|\omega_n\rangle$ is ed at site n for almost every n and, the localization lengths being small enough for ν, we can effectively replace $|u_n\rangle$'s by $|\omega_n\rangle$'s in (29). Since $\hat{H}_0$ is diagonal in the s, the Hamiltonian with the kicking can now be written as

$$\hat{H} = \sum_n e_n |\omega_n\rangle\langle\omega_n| + \sum_{n,s} v(|\omega_n\rangle\langle\omega_{n+1}| + |\omega_n\rangle\langle\omega_{n-1}|)\delta(t - sT). \tag{30}$$

choice of the sequence of eigenvalues e_n and the replacement of the $| u_n\rangle$'s by the 's, based on the results of § 2, is crucial to the scheme being considered here. It ies that we are in the insulator regime $\nu > \nu_c$, where we can now employ the land construction as in the case of (17). In the present situation $\hat{V}$ does possess est neighbour matrix elements only, and the matrix element is independent of the site e energy lattice. Since $\tilde{V}$ is a regular function of $\hat{V}$, we can assume similar properties $_{nm}$ as well. What is essential here is that $\tilde{V}_{nm}$ should not vary irregularly with n and nd that it should not have significantly large values for large $| n - m |$, both of which be seen to follow as consequences of (30). In fact, to a good degree of oximation, we arrive at the equivalent tangent potential model

$$c_n u_n + \tilde{\lambda}(u_{n+1} + u_{n-1}) = 0, \tag{31}$$

re

$$\tilde{\lambda} = -\frac{1}{2\pi} \int_0^{2\pi} \mathrm{d}\theta \cos\theta \tan\left(\frac{v}{2}\cos\theta\right), \tag{32}$$

the on-site tangent potentials are given by (24b), with E_n's replaced by the lattice red e_n's. The binding-to-hopping ratio in this equivalent TBM is $\tilde{\lambda}^{-1}$ which decreases increasing v. For small enough T the sequence $\{c_n\}$ is essentially similar to $\{e_n\}$ and dicated in § 2, this in turn is similar to $\{V_n\}$. Hence, for small T, a *m-i type transition rs in the tangent potential TBM in an essentially similar manner as in the original as v, the strength of kicking, is made to vary*. In particular, for small enough v, all the nstates in this equivalent lattice are localised while, for larger v, extended states inate the spectrum. This means that for small v, the quasistationary states of the time

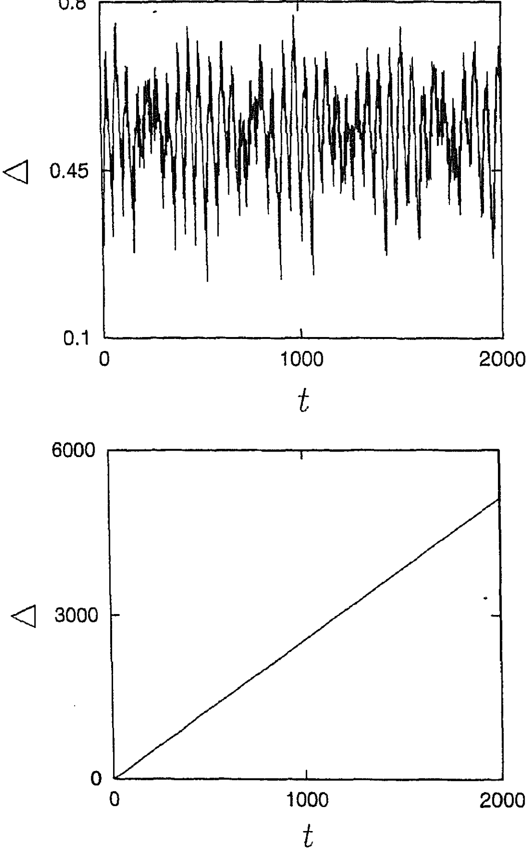


Figure 6. Root mean squared spread (Δ) of wavepacket in the energy lattice against number of kicks (t) for the kicked FL for $\nu = 7.0$ with kicking parameters (**a**) $T = 0.1$, $v = 0.1$ and (**b**) $T = 0.1, v = 2.0$. In each case the initial wavefunction was assumed to be concentrated at the centre of the lattice and the quantum map of (34) was numerically iterated, retaining Bessel function coefficients J_k up to $|k| = 6$. The quasiperiodic nature of the time evolution in (a) is evident while in (b) the spread of the wavefunction appears to be ballistic.

dependent problem (28) are localised in the energy lattice while, for larger v, there appear quasistationary states extended in the energy lattice.

This enables us to make predictions, in the context of the periodically driven Hamiltonian (28), about the time evolution of wavepackets *in the energy lattice*. For instance, for sufficiently small T and large enough v, the presence of extended quasi-energy states with associated absolutely continuous parts in the spectrum implies that there occurs a ballistic spread of wavepackets in the energy lattice, while for smaller values of v one expects a recurrent time evolution.

The time evolution of a wavepacket

$$|\psi\rangle = \sum_n p_n(t)|\omega_n\rangle, \tag{33}$$

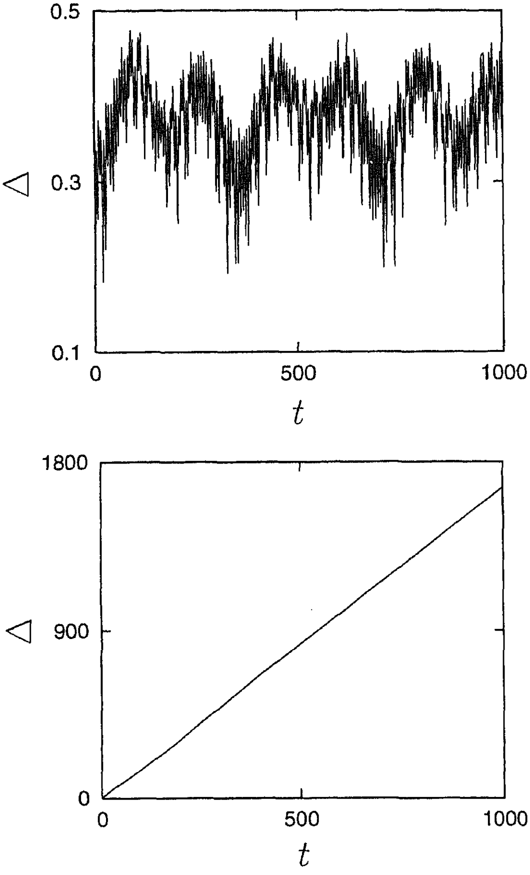


Figure 7. Root mean squared spread (Δ) of wavepacket in the energy lattice against number of kicks (t) for the kicked SEL for $\nu = 6.0$ with kicking parameters **(a)** $T = 0.1$, $v = 0.1$, and **(b)** $T = 0.1, v = 2.0$. See legend to figure 6.

governed by the Hamiltonian (30) is given by the quantum map

$$p_n^{s+1} = \sum_k \exp\left[-i\left(e_{n+k}T + \frac{\pi k}{2}\right)\right] J_k(2v) p_{n+k}^s, \tag{34}$$

where $p_n^s \equiv \lim_{\epsilon \to 0^+}(sT + \epsilon)$ and J_k is the Bessel function of order k. This map can be iterated numerically for given v, T. For the sake of convenience we take for $s = 0$

$$p_n^0 = \delta_{mn}$$

for some given m (i.e., an eigenstate $|\omega_m\rangle$ as the initial $|\psi\rangle$), and on iterating through s impulses compute

$$\bar{n}(s) \equiv \sum_n n|p_n^s|^2 \tag{35a}$$

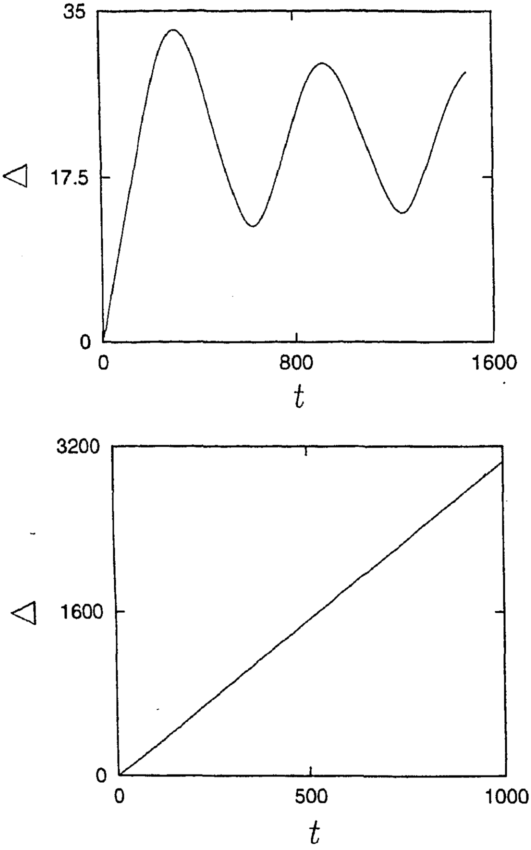


Figure 8. Root mean squared spread (Δ) of wavepacket in the energy lattice against number of kicks (t) for the kicked CSL for $\nu = 6.0$ with kicking parameters **(a)** $T = 0.1$, $v = 0.1$, and **(b)** $T = 0.1, v = 2.0$. See legend to figure 6.

and

$$\Delta^2(s) \equiv \sum_n (n - \bar{n}(s))^2 |p_n^s|^2, \tag{35b}$$

which are respectively the mean position and mean squared spread of the wavepacket *on the energy lattice* after s impulses. While the right hand side of (34) involves an infinite summation, in practice one can truncate after a few terms since the Bessel functions decrease with increasing $|k|$.

Figures 6(a,b) present results of numerical iteration of the quantum map (34) for the Feigenbaum lattice (with $\nu > \nu_c$) for two different sets of driving parameters. The agreement with the predictions of the above theory is immediately apparent. Thus, as the kicking strength v is increased from 0.1 to 2.0, the quasiperiodic fluctuations in the mean squared spread of the wavepacket are seen to give way to a unbounded ballistic growth.

Similar results are encountered with for the SEL and the CSL as well, as seen from figures 7(a,b) and 8(a,b) respectively. In other words, the theory outlined above does indeed give us correct predictions about the quasienergy spectra and wavepacket dynamics in the energy lattice for the TBM's with 'slowly varying' on-site potentials being considered in this paper. It should be mentioned that motion in the energy lattice is not quite the same thing as that in the real space lattice, and so the findings obtained here cannot be immediately compared with the corresponding ones on the KHM. Nevertheless the above theory is evidently useful in leading to a formulation of criteria for recurrent and not-recurrent dynamics for at least a class of mesoscopic systems. Since the theory does not refer to the classical phase space, these may be looked upon as purely quantum criteria independent of criteria for classical chaos. Such criteria appear to be necessary in the context of findings on the kicked Harper model.

[*Remark.* It is neccessary at this point to set right a computational error in [20, 21] where numerical iteration of the quantum map for the CL and the QPL in presence of kicking appeared to give rise to normal diffusion rather than ballistic spread for large values of kicking strength. This, however, was an artefact caused by the numerical programme which retained too few terms in the right hand side of (34), thus truncating away a number of terms with significant values. In the present series of results, on the other hand, we have retained Bessel functions J_k with $|k| \leq k_{\max}$, where $k_{\max}$ has been chosen sufficiently large. As indicated above, a ballistic spread for large v is actually more in keeping with the prediction from the theory we have outlined.]

4. AC Conductivity of incommensurate lattices

The a.c electrical conductivity $\sigma(\omega)$ at low temperatures (T) in incommensurate lattices is a good probe for the nature of spectra and wavefunctions. While the d.c limit gives us only local information on individual states near the Fermi level, the ω-dependence of the a.c conductivity is a vastly richer source of information, involving the *global* structure of the spectrum and of the wavefunctions. Even the small ω regime at $T \to 0$ gives us new insights, especially concerning extremely distant but resonating localized states. The ω-dependence in this regime already shows subtle variations with the arithmetic nature of the incommensuration, and differs from corresponding results on completely disordered lattices [57]. From the applied point of view, finite-ω a.c conductivity in incommensurate lattices provides the context for developing materials with interesting frequency-selective transport properties, differing from ordered structures on the one hand, and from completely disordered ones on the other.

However, precisely because of the fact that the a.c conductivity depends on the global structure of the spectrum and of the wavefunctions, its theoretical calculation or numerical computation is beset with formidable difficulties. In particular, the knowledge of *all* the stationary state wavefunctions, which is necessary for the computation of $\sigma(\omega)$, is a severe pre-requisite for large systems. The principal difficulty lies in the fact, indicated in § 2, that wavefunctions vary extremely erratically with energy. Thus, as seen from figure 3, wavefunctions with energies close to one another have their peaks at widely different locations.

In [22], the numerical observations of § 2 were used in the context of the CL and the QPL to establish that *gross features of the ω-dependence of the a.c conductivity can be accounted for quite satisfactorily in terms of the energy spectra alone, without direct reference to the wavefunctions.*

A simple model based on the correlations outlined in § 2 was proposed for the computation of the gross features in the variation of $\sigma(\omega)$ and was found to compare favourably with numerical computations of $\sigma(\omega)$ in accordance with the Kubo-Greenwood formula.

Since we have seen that these correlations hold for other incommensurate lattices admitting of m-i type transitions, in particular the FL, the SEL and the CSL, the above simple model should in principle work for these lattices as well.

In this section we summarise relevant background from [22] and then show that the domain of applicability of the simple model does indeed include these other lattices as well. We point out in the next section a few features of $\sigma(\omega)$ for the lattices considered, by way of concluding remarks.

4.1 *Resonance peaks in a.c conductivity*

At $T \to 0$ there are no phonons in a lattice assisting the conductivity and, in the single particle approximation, the a.c conductivity as a function of the frequency ω is given (up to constants, in units of $e^2/\hbar$) by the Kubo–Greenwood formula (see, e.g., [16])

$$\sigma(\omega) = \frac{\omega}{\Omega} \sum_{m,n} | \hat{x}_{mn} |^2 (f_n - f_m) \delta(E_m - E_n - \omega), \tag{36}$$

which has been written in a form suitable for a system with localised states. Here Ω stands for the system size, the E_m's are the energy eigenvalues, the f_m's are the Fermi occupation probabilities, and $\hat{x}_{mn}$ stands for the matrix element

$$\hat{x}_{mn} = \sum_{l} l \psi_l^{(m)} \psi_l^{(n)}, \tag{37}$$

where $\psi_l^{(m)}$ is the component at lattice site l of the normalised wavefunction corresponding to energy E_m. In (36), the indices m, n appear under the summation, and hence the order in which the energy eigenvalues and the corresponding eigenfunctions are arranged does not matter.

Equation (36) tells us that, apart from an over-all factor ω, the variation of σ with ω involves a series of resonance peaks, a peak at any given frequency ω arising when there exists a pair of states, one below and the other above the Fermi level, such that the energy interval between the two is equal to ω (in units of $\hbar$) and the co-ordinate matrix element between the two states is appreciable. Such a pair of states may be termed a 'resonating' pair. It may be noted that, in general, a resonance peak in $\sigma(\omega)$ may involve more than one pair of resonating states. Indeed, experimentally it is not $\sigma(\omega)$, but an average taken over some small ω-interval that is usually relevant. In this averaged σ, each resonance peak will more often than not actually involve more than one pair of resonating states. Since all states are involved in the summation (36), and since there may be resonating states distributed throughout the spectrum, it is apparent that the calculation of $\sigma(\omega)$ requires the knowledge of the entire spectrum and the complete set of wavefunctions.

Early computations of low temperature a.c conductivity for 1D random lattices were by Penchina and Mitchell [58], and by Albers and Gubernatis [59] (see [17]). The resonances are dense in a random lattice, with strong fluctuations in the finite ω regime. Mott (see, e.g., [60]) argued that, in the small ω regime, σ has a ω^2 behaviour, with a logarithmic ω-dependence thrown in due to the so-called multiplet formation by extremely distant resonating states. Prange, Grempel, and Fishman [57] calculated $\sigma(\omega)$ in an exactly solvable incommensurate TBM, and pointed out that numerous possible *types* of behaviour in the small ω regime might be possible for different types of incommensuration, which are clearly distinct from the ω dependence of a random lattice.

However, large ω regimes are largely outside the purview of these earlier studies, since no systematic computation of the resonances could be possible without a knowledge of all the wavefunctions $\psi^{(m)}$ and since, as already seen in the small ω regime, resonance involves apparently uncorrelated states for random and quasirandom lattices.

4.2 *Gross structure of $\sigma(\omega)$: Locating the resonant peaks*

In this context the results presented in § 2 shed a new light on the problem. Indeed, the calculation of $\sigma(\omega)$ becomes immediately simpler as soon as one takes the sequence $\{E_n\}$ to be the lattice-ordered sequence of energies since we already know that in this ordering the gross structures of the corresponding wavefunctions are strongly correlated with the energies themselves. This enables us to compute the locations of a considerable fraction of the resonant peaks from a knowledge of the spectrum alone, using the lattice-ordered sequence of energies to supply the relevant gross features of the wavefunctions. What cannot be captured in this approach is the fine structure in the variation of σ, namely the actual sizes of these resonant peaks, since this does in fact demand a detailed knowledge of the wavefunctions.

Thus, we make use of the fact that when the energies are reckoned in the lattice-ordered sequence, the wavefunctions $\psi^{(m)}$ and $\psi^{(n)}$ involved in the matrix element $\hat{x}_{mn}$ are peaked at lattice sites m, n respectively and this matrix element can have a substantial magnitude only when the wavefunctions have a considerable overlap. Remembering that the wavefunctions themselves are exponentially localized with certain localization widths, it is apparent that the matrix element $\hat{x}_{mn}$ can be appreciable *only when the combined widths of the wavefunctions involved is less than the distance between their peaks.* In this context we use the Thouless formula [61] for the inverse localization lengths in a 1D chain:

$$\gamma_m = \sum_{n \neq m} \ln \mid (E_n - E_m)/\lambda \mid . \tag{38}$$

This formula relates the inverse localization lengths with the spectrum. So, the spectrum alone supplies all the information required in the calculation of the gross structure of $\sigma(\omega)$ if we adopt the following working hypothesis: *two states $\psi^{(m)}$ and $\psi^{(n)}$ at an energy gap ω constitute a pair of resonating states only if $\mid m - n \mid$ is less than ${\gamma_m}^{-1} + {\gamma_n}^{-1}$.*

This working hypothesis implies that the matrix element $\hat{x}_{mn}$ is appreciable only when $\mid m - n \mid$ is small enough, and is based on the correlation between the peaks of the wavefunctions and the energies in the lattice-ordered sequence. As such, it is expected to hold for the incommensurate TBM's we are considering, and *not* for a random lattice. This was corroborated numerically for the CL and QPL *vis a vis* the random lattice in ref. [22]. In figure 9(a,b) we plot the magnitude of the matrix element, calculated for a fixed

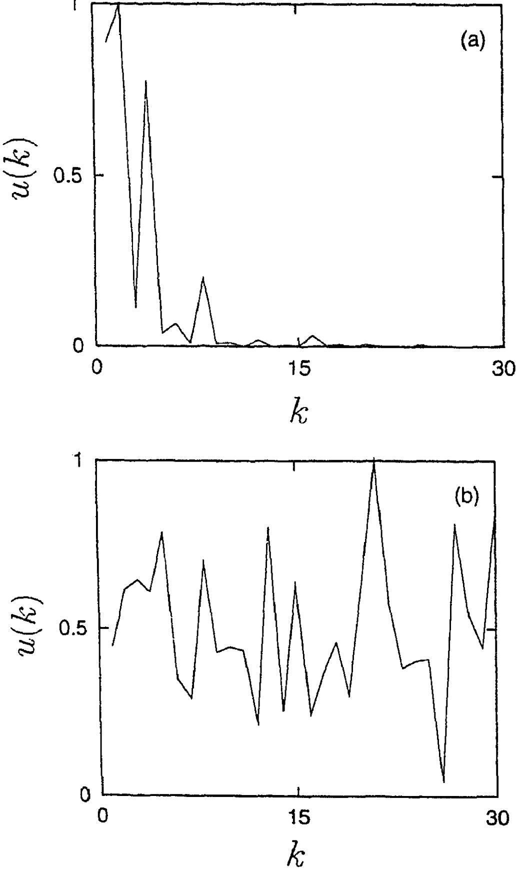


Figure 9. Average overlap $u(k) \equiv N^{-1} \sum_m | \tilde{x}_{m,m+k} |$, N = lattice size (for notations, see text), against separation k between peaks for (**a**) the FL with $\nu = 7.0$, and (**b**) a random lattice with $\nu = 5.0$. In the plots, max $u(k)$ has been normalised to 1.

separation between the peaks and *averaged* over the lattice, as a function of the separation for the FL and the random lattice respectively. It is apparent from these plots that the above working hypothesis is indeed tenable for the FL while, for the random lattice, the matrix element fluctuates without any appreciable decimation as the distance between the peaks is made to increase.

With this justification for our working hypothesis and following [22], we compute the locations of the resonances as follows: in the Kubo formula (36), we substitute $\tilde{x}_{mn} = 1$ whenever $| m - n | < {\gamma_m}^{-1} + {\gamma_n}^{-1}$, and zero otherwise. We do not distinguish between the magnitudes of the non-zero matrix elements since we are interested primarily in the *locations* of the resonances.

Numerical results of [22] indicate that this approach, simple as it is, is surprisingly successful in reproducing correctly the gross features of the variation of σ with ω for the CL and the QPL.

We present below corresponding numerical evidence for the other TBM's introduced in this paper, namely the FL, the SEL, and the CSL.

Variations of the zero temperature a.c conductivity with frequency, as computed from the Kubo formula (36) with the help of numerically computed wavefunctions, are shown in figures 10(a,b,c) for the FL, the SEL and the CSL respectively. One finds that while the resonant peaks are clustered in band-like groups for the latter two lattices, they are relatively sparsely distributed in the self-similar Feigenbaum lattice. In this respect the FL resembles the CL since, in both these lattices, nearby energies in the lattice ordered sequence are clustered together into closely packed groups along the energy scale.

In order to compare the gross features of this variation, computed from the Kubo formula, with the results of our above simplified computation of the resonance peaks, we first ignore those resonant peaks obtained from the Kubo formula whose magnitudes are less than some small suitably chosen fraction (r_1) of the most pronounced peak, thereby eliminating the minor peaks from the comparison, and then set $\sigma = 1$ for all the remaining peaks resulting from the Kubo formula and display them as upward spikes in a $\sigma - \omega$ plot. While computing these resonant peaks, the relevant ω-interval (depending on the width of the spectrum in the energy scale) was sampled by a number of points

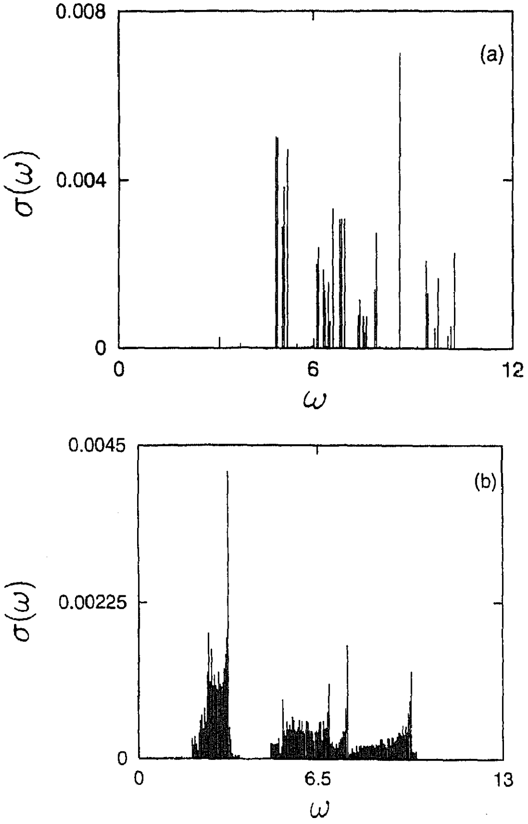


Figure 10a-b. (*Continued*)

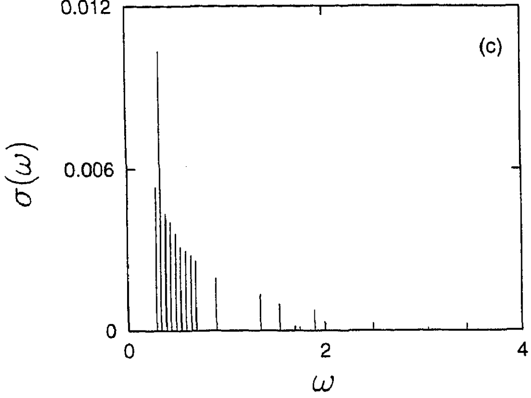


Figure 10. Plot of $\sigma(\omega)$ vs ω for (**a**) the FL with $\nu = 7.0$, (**b**) the SEL with $\nu = 6.0$, and (**c**) the CSL with $\nu = 4.0$. Each spike represents an average of σ for a small interval (0.025) of frequency; data points are at ω-intervals 0.05. Fermi levels, to which the plots are not particularly sensitive, were chosen at energies $0.7509\cdots$, $-1.3505\cdots$, and $-0.0815\cdots$ respectively.

distributed uniformly, and at each sampling point a *coarse-grained* σ was calculated by including all pairs of resonating states whose energy separation lies within an interval ω to $\omega + \delta\omega$, for some suitably chosen small $\delta\omega$. Likewise, the resonances were computed from the simplified model for the same set of sampling points and with the same coarse-graining, and now σ is set at -1 at all these resonances, superposing these in the same $\sigma - \omega$ plot as downward spikes. Analogous to the use of the fraction r_1 in eliminating the relatively minor upward spikes, a fraction r_2 has been used for the downward spikes and, in some instances, r_2 has been chosen different from r_1 for facility of comparison. In chosing this mode of presentation, we ignore the variations in the magnitude of the a.c conductivity at the resonant peaks, instead concentrating on the locations of the more major peaks.

Figures 11, 12, and 13 are results for the FL, the SEL and the CSL respectively. One finds in each case that, considering the simplicity of the model, a surprisingly large fraction of actual resonances are reproduced without explicit use of the wavefunctions and what is more, there appears to be quite a substantial agreement even for values of ν *below* the limiting value ν_c where extended states exist in the spectrum dominated by localized states. Even though the correlation expressed by (8) does not make sense for all the wavefunctions for $\nu < \nu_c$, till it does continue to hold good for a considerable fraction of the localized states; and the contribution of these states to $\sigma(\omega)$ makes our simplified model useful even before the m-i transition is completed.

For instance, the coarse-grained σ's computed in the two approaches agree at a fraction 0.926 of the total number of sampling points for the FL with $\nu = 5.0$ (figure 11(a); most of these data points correspond to $\sigma = 0$), and 95.4% of the upward spikes are actually reproduced by the downward ones. As expected, the agreement improves as ν is made to increase beyond the limiting value into the insulator regime. Thus, for the FL with $\nu = 7.0$ (figure (13(b)), a fraction 0.99 of the coarse-grained σ's calculated in the two approaches agree (most of which, once again, correspond to $\sigma = 0$) and *all* the upward

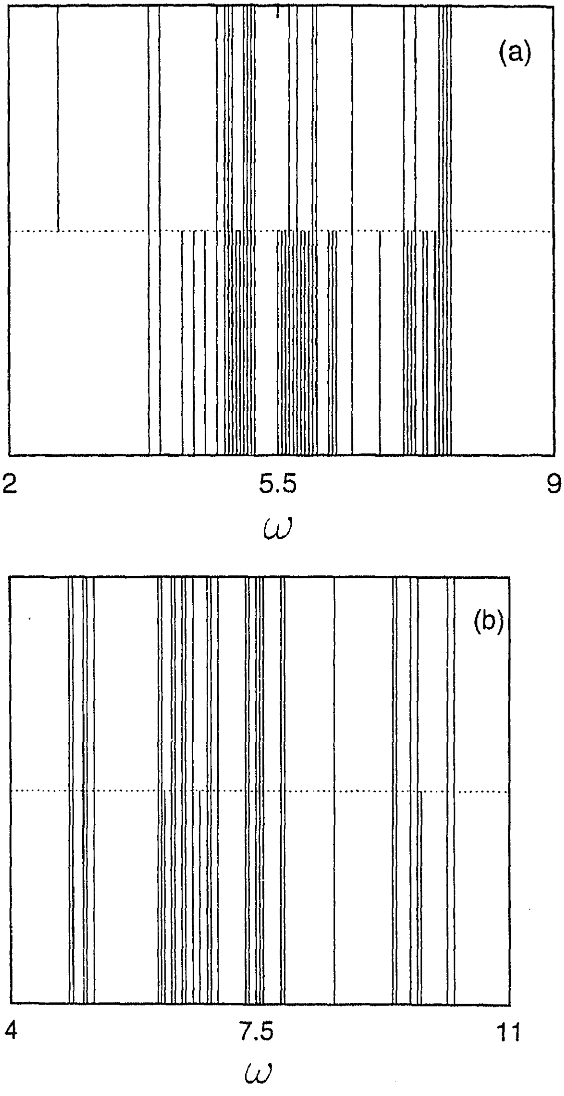
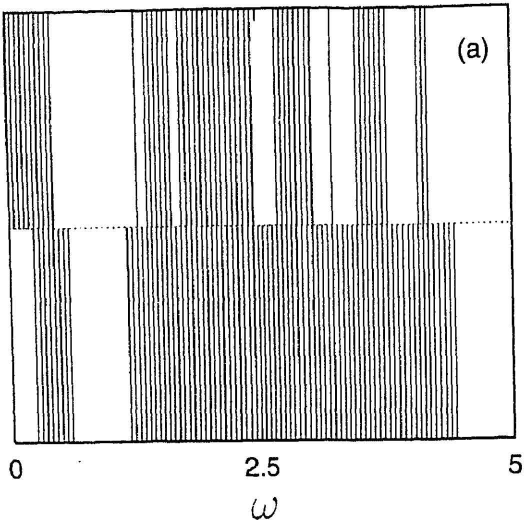


Figure 11. Locations of major resonance peaks in a.c conductivity in accordance with Kubo formula with numerically obtained wavefunctions (upward spikes) compared with the locations obtained from the simplified model of the overlap $\hat{x}_{mn}$ (downward spikes) as explained in the text, for the FL. Notations : r_1, r_2 as explained in the text; ϵ_F, Fermi level. (**a**) $\nu = 5.0, r_1 = 0.02$, $r_2 = 0.02, \epsilon_F = 0.3364 \cdots$, (**b**) $\nu = 7.0$, $r_1 = 0.04, r_2 = 0.01$, $\epsilon_F = 0.7509 \cdots$. Calculated values of σ have been averaged over ω-intervals of width 0.025, and data points are at ω-intervals 0.05 as in figure 10.

spikes are reproduced by downward ones, even while our model yields some downward spikes of its own, not actually there according to the Kubo formula. Similarly, for the Soukoulis–Economou model, the two computations agree at a fraction 0.893 of the sampling points, and 92% of upward spikes are reproduced by downward ones even at



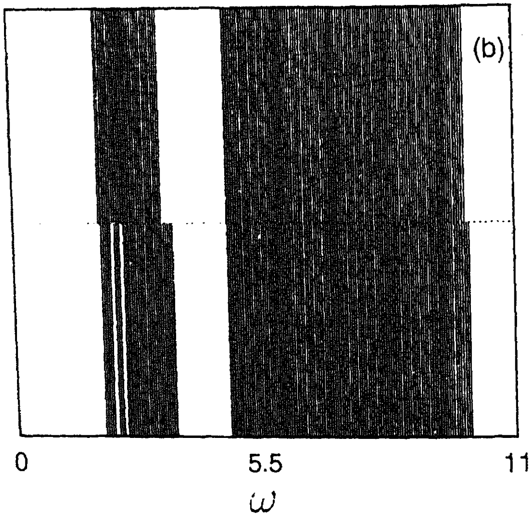


Figure 12. Comparison of locations of peaks in a.c conductivity, for the SEL (see legend to figure 11; same notations). (**a**) $\nu = 1.5$, $r_1 = 0.03$, $r_2 = 0.01$, $\epsilon_F = -0.145\cdots$, (**b**) $\nu = 6.0$, $r_1 = 0.02, r_2 = 0.02, \epsilon_F = -1.3505\cdots$.

$\nu = 1.5$ (figure 12(a)), while for $\nu = 6.0$ (figure 12(b)) the corresponding figures are 0.966 and 97.7% respectively. The agreement is comparatively poorer for the cosine model, where the σ's computed from the two approaches agree at a fraction 0.815 of the data points and 77.7% of the upward spikes are reproduced in the set of downward ones (figure 13(a)) for $\nu = 1.5$ which is less than the limiting value. The corresponding figures for $\nu = 4.0$-well inside the insulator regime – are 0.985 and 84.6% respectively (figure 13(b)). Moreover, in the CSL with $\nu < \nu_c$ (figure (13(a)), our simple model appears to pick up numerous spurious resonant peaks at the high frequency end of the spectrum.

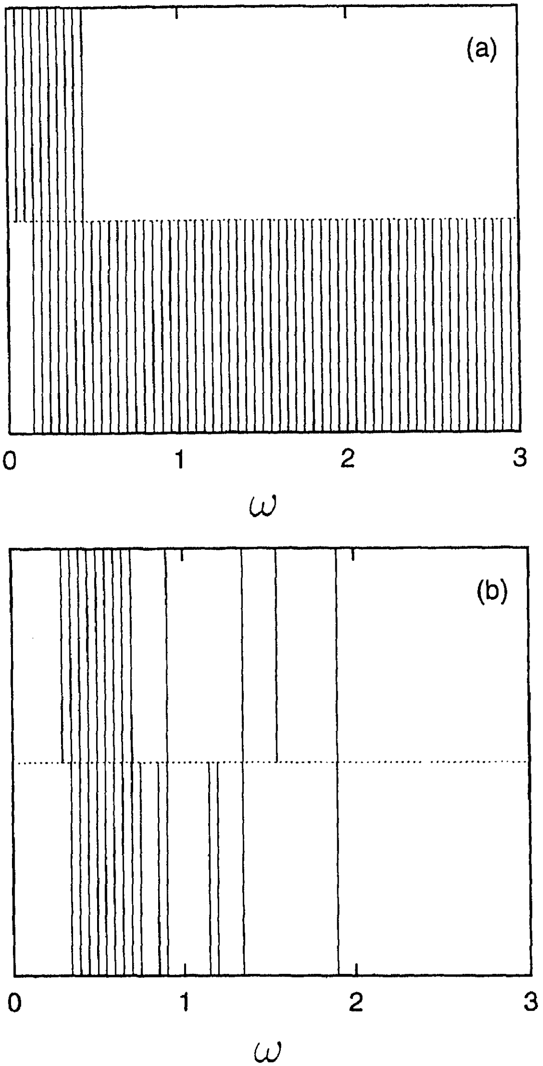


Figure 13. Comparison of locations of peaks in a.c conductivity, for the CSL (see legend to figure 11; same notations). (**a**) $\nu = 1.5, r_1 = 0.03, r_2 = 0.01$, $\epsilon_F = -0.2308\cdots$, (**b**) $\nu = 4.0$, $r_1 = 0.02, r_2 = 0.02$, $\epsilon_F = -0.0815\cdots$.

This discernible discrepancy in the CSL means that the perturbation series for the energies and wavefunctions in the perturbation scheme indicated in § 2 does not work so well in this case. However, even here the level of agreement is not one to be dismissed lightly.

5. Summary and concluding remarks

In this paper we have reviewed a set of known results on unusual spectral properties and related features of wavefunctions in 1D incommensurate TBM's as well as on features relating to recurrent and non-recurrent quantum dynamics in these systems. We have presented numerical results on a range of questions relating to a set of TBM's characterised by what may be looked upon as 'slowly varying' on-site potentials and by m-i type transitions in their spectra. These results supplement corresponding findings on a couple of TBM's of the above type, presented in a series of earlier papers. All these results are based on a set of interesting correlations concerning their spectra and wavefunctions that are seen to hold in the localized regime, beyond the limiting value of the binding-to-hopping ratio, and we have presented a qualitative explanation of these correlations on the basis of perturbation theory.

We have also reviewed a set of basic results on the quantum dynamics of wavepackets in mesoscopic systems and, in this context, have made use of the above set of correlations for the TBM's under study to arrive at a model, based on the so-called Maryland construction, that enables us to compute the quasienergy spectra for problems involving periodic kicking in these lattices. While we have not actually computed the quasienergy spectra due to the prohibitive size of the computations involved, we have tested qualitative predictions from the model against numerical iterations of the relevant quantum maps.

While our model of wavepacket dynamics can be used to compute the quasienergy spectra from the equivalent stationary TBM involving the 'tangent potential', it is to be noted that the dynamics of wavepackets on this equivalent TBM is *not* the same as wavepacket dynamics in the original periodically driven problem that we start with because the equivalent TBM is after all nothing more than an artefact. Hence, results on correlation between diffusion in a stationary TBM and the corresponding spectral characteristics are not directly relevant in this problem. On the other hand, features of the wavefunctions in the equivalent TBM's such as whether these are extended or localised are, to a certain extent relevant and we have made use of these features to arrive at the qualitative predictions which do agree with results of numerical iterations on the quantum maps. However, the types of wavepacket spreading we have encountered in our numerical iterations cannot be said to be exhaustive in the sense of being comparable in variety to those found in, say, the kicked Harper's model. For instance, we have not encountered the diffusive spread, either normal or 'anomalous', of the wavepackets. It is not immediately clear whether this is due to the fact that the wavepacket dynamics we have looked into is in the energy lattice rather than in the real space lattice. A more precise and thorough analysis of the wavepacket dynamics in the context of our model will be presented in a future communication.

Closely related to the dynamics of wavepackets are the transport properties of the lattices. Transport in 1D incommensurate lattices is characterized by numerous novel features due to exotic spectral characteristics. Conversely, experimentally observed features of transport in pseudorandom lattices can be used as probes to gain information on spectral features and wavefunctions. In particular, a.c conductivity at low temperatures is a highly sensitive probe in this respect as revealed from early studies in the low frequency regime. However, the theoretical study or numerical computation of the a.c conductivity in the finite frequency regime is beset with formidable difficulties because it requires a knowledge of the entire spectrum and the complete set of wavefunctions. We have made use of the above correlations established for the TBM's under study to arrive at a simple model for predicting the gross structure of the variation of the a.c conductivity with frequency, and have shown that this simple model does indeed predict the gross structure quite successfully.

The a.c conductivity of incommensurate lattices appears to be of special interest since the variation of a.c conductivity with frequency, consisting of series of resonant peaks, is of a selective nature and is moreover highly specific to details of structure of the lattice, unlike the dense fluctuations one observes in random lattices or the featureless variations in periodic lattices. This is apparent from figures 10(a,b,c) where the locations of the resonant peaks in $\sigma(\omega)$ are seen to follow distinct patterns for the FL, the SEL, and the CSL. For instance, most of the resonances fall into band-like groups in the SEL and the CSL, while they are less clustered in the FL, even though there appears to be points of accumulation on the frequency scale in this case. In this context, our simple model leading to considerably successful predictions on the locations of these resonant peaks acquires significance.

References

[1] D Schechtman, I Blech, D Gratias and J W Cahn, *Phys. Rev. Lett.* **53**, 1951 (1984)
[2] D Weiss, M L Roukes, A Menschig, P Grambow, R von Klitzing and G Weimann, *Phys. Rev. Lett.* **66**, 2790 (1991)
[3] R Merlin, K Bajema, R Clarke, F Y Jung and P K Bhattacharya, *Phys. Rev. Lett.* **55**, 1768 (1985)
[4] F Axel and H Terauchi, *Phys. Rev. Lett.* **66**, 2223 (1991)
[5] D R Hofstadter, *Phys. Rev.* **B14**, 2239 (1976)
[6] S Aubry and G Andre, *Proc. Israel Phys. Soc.* edited by C G Kuper **3**, 133 (1979)
[7] M Kohmoto, L P Kadanoff and C Tang, *Phys. Rev. Lett.* **50**, 1870 (1983)
[8] S Ostlund, R Pandit, D Rand, H J Schellnhuber and Siggia, *Phys. Rev. Lett.* **50**, 1873 (1983)
[9] G Casati, B V Chorikov, F M Izrailev and J Ford, in *Lecture notes in physics* edited by G Casati and J Ford (Springer, Berlin, 1979) vol. 93, 334
[10] B V Chirikov, F M Izrailev and D L Shepelyansky, *Physica* **D33**, 77 (1988)
[11] R Artuso, F Borgonovi, I Guarneri, L Rebuzzini and G Casati, *Phys. Rev. Lett.* **69**, 3302 (1992)
[12] F Craciun, A Bettucci, E Molinari, A Petri and A Alippi, *Phys. Rev. Lett.* **68**, 1555 (1992)
[13] A Petri, A Alippi, A Bettucci, F Craciun and F Farrelly, *Phys. Rev.* **B49**, 15067 (1994)
[14] J B Sokoloff, *Phys. Rep.* **126**, 129 (1985)
[15] D Wurtz, T Schneider, A Politi and M Zannetti, *Phys. Rev.* **B39**, 7829 (1989)
[16] D J Thouless, *Phys. Rep.* **13**, 93 (1974)
[17] W M Visscher and J E Gubernatis, in *Dynamical properties in solids* edited by G K Horton and A A Maradudin, (North Holland, Amsterdam, 1980) vol. 4
[18] T K Roy and A Lahiri, *J. Phys. Condens. Matter* **3**, 1153 (1991)
[19] A Lahiri, *Phys. Rev. Lett.* **70**, 1611 (1993)
[20] A Lahiri, in *Computational aspects in chaos and nonlinear dynamics* edited by G Ambika and V M Nandakumaran (Wiley Eastern Ltd., New Delhi, 1994) p 107
[21] A Lahiri, *Phys. Rev.* **B49**, 14925 (1994)
[22] A Lahiri, *Phys. Rev.* **B53**, 3702 (1995)
[23] M J Feigenbaum, *J. Stat. Phys.* **19**, 25 (1978); **21**, 669 (1979)
[24] N W Ashcroft and N D Mermin, *Solid state physics* (Saunders College, Philadelphia, 1976) ch. 10
[25] J M Ziman, *Models of disorder, the theoretical physics of homogeneously disordered systems* (Cambridge University Press, Cambridge, 1979) ch. 8
[26] K Ishii, *Prog. Theor. Phys. Suppl.* **53**, 77 (1973)
[27] J Hori, *Spectral properties of disordered chains and lattices* (Pergamon, Oxford, 1968) ch. 5
[28] M Reed and B Simon, *Methods of modern mathematical physics* (Academic Press, N.Y. 1978) vol. 4, ch. 13

[30] J Avron and B Simon, *Bull. Am. Math. Soc.* **6**, 81 (1982)
[31] J Bellisard and B Simon, *J. Funct. Anal.* **48**, 408 (1982)
[32] C M Soukoulis and E N Economou, *Phys. Rev. Lett.* **48**, 1043 (1982)
[33] A Chakrabarti, *J. Phys. Condens. Matter* **8**, L99 (1996)
[34] S Das Sharma, S He and X E Xie, *Phys. Rev. Lett.* **61**, 2144 (1988)
[35] M Griniasty and S Fishman, *Phys. Rev. Lett.* **60**, 1334 (1988)
[36] N Brenner and S Fishman, *Nonlinearity* **4**, 211 (1992)
[37] A Crisanty, *J. Phys.* **A22**, L681 (1989)
[38] J J Sakurai, *Modern quantum mechanics* (Addison-Weseley Publishing Co., Reading, Massachusettes, 1994) ch. 5
[39] A Heslot, *Phys. Rev.* **D31**, 1314 (1985)
[40] B Eckhardt, *Phys. Rep.* **163**, 206 (1988)
[41] I Guarneri, *Europhys. Lett.* **10**, 95 (1989)
[42] I Guarneri, *Europhys. Lett.* **21**, 729 (1993)
[43] R Ketzmerick, G Petschel and T Geisel, *Phys. Rev. Lett.* **66**, 1651 (1991)
[44] N Hiramoto and S Abe, *J. Phys. Soc. Jpn.* **57**, 1365 (1988)
[45] T Geisel, *Chaos and Quantum Phenomena in Semiconductor Microstructures*, Notes of Lecture presented at Mini-Workshop on Nonlinearity (26 July–6 August 1993) International Centre for Theoretical Physics, Trieste, Italy
[46] D E Katsanos, S N Evangelou and S J Xiong, *Phys. Rev.* **B51**, 895 (1995)
[47] J Avron and B Simon, *J. Funct. Anal.* **43**, 1 (1981)
[48] T Geisel, J Wagenhuber, P Niebauer and G Obermeir, *Phys. Rev. Lett.* **64**, 1581 (1990)
[49] J Wagenhuber and T Geisel, *Phys. Rev.* **B45**, 4372 (1992)
[50] R Lima and D Shepelyanski, *Phys. Rev. Lett.* **67**, 1377 (1991)
R Ketzmerick, G Petschel and T Geisel, *Phys. Rev. Lett.* **67**, 3635 (1991)
[51] T Hogg and B A Huberman, *Phys. Rev. Lett.* **48**, 711 (1982)
[52] A Peres, *Phys. Rev. Lett.* **49**, 1118 (1982)
[53] J Bellisard, in *Lecture notes in mathematical physics* (Springer, Berlin, 1985) vol. 1049
[54] R Artuso, G Casati and D Shepelyanski, *Phys. Rev. Lett.* **68**, 3826 (1992)
[55] S Fishman, D R Grempel and R E Prange, *Phys. Rev. Lett.* **49**, 509 (1982)
[56] D R Grempel, R E Prange and S Fishman, Phys. Rev. **A29**, 1639 (1984)
[57] R E Prange, D R Grempel and S Fishman, *Phys. Rev. Lett.* **53**, 1582 (1984)
[58] C M Penchina and D L Mitchell, *J. Non-Cryst. Solids* **7**, 127 (1972)
[59] R Albers and J E Gubernatis, *Phys. Rev.* **B17**, 4487 (1977)
[60] N F Mott and E A Davis, *Electronic processes in non-crystalline materials* (Clarendron, Oxford, 1971)
[61] D J Thouless, *J. Phys.* **C5**, 77 (1972)

Fractal characteristics of critical and localized states in incommensurate quantum systems

INDUBALA I SATIJA and JUKKA A KETOJA*
Department of Physics and Institute of Computational Sciences and Informatics, George Mason University, Fairfax, VA 22030, USA
*Department of Mathematics, P.O. Box 4, FIN-00014, University of Helsinki, Finland

Abstract. Incommensurate quantum systems with two competing periodicities exhibit metallic (with Bloch-type extended wave functions), insulating (with exponentially localized wave functions) as well as *critical* (with fractal wave functions) phases. An exact renormalization method, which takes into account the inherent incommensurability, is used to obtain the phase diagram of various quantum models for the two-dimensional electron gas and for quantum spin chains in a magnetic field. In this approach, the scaling properties of the fractal eigenstates are characterized by a fixed point or a strange invariant set of the renormalization flow. One of our novel results is the existence of self-similar fluctuations in the localized states once the exponentially decaying envelope is factorized out. In almost all cases under investigation here, the universality classes can be broadly classified as those of the nearest-neighbor square or triangular lattices.

Keywords. Incommensurate systems; localization; fractal.

PACS Nos 72.15; 64.60

1. Introduction

Incommensurate or quasiperiodic (QP) systems with two irrationally related frequencies provide a class of aperiodic models which have been of great interest to experimentalists, theoreticians as well as mathematicians. Incommensurate structures appear in many physical systems such as quasicrystals, two-dimensional electron systems, magnetic superlattices, spin and charge-density waves as well as in one-dimensional ionic conductors. [1] The mathematical interest in QP systems is due to the small-divisor problems [2] which arise in that context. Theoretically, the subject is of great interest because these systems are considered to be in between periodic and random. Periodic systems exhibit Bloch-type extended (E) states and a continuous spectrum while for a random potential all states are localized (L) in one dimension. However, QP systems exhibit the metal-insulator (or E-L) transition in one dimension [1]. At the transition point the eigenfuntions are neither extended nor exponentially localized and are referred to as *critical* (C). The spectrum in QP systems can be either absolutely continuous, point-like, singular continuous, or a mixture of these [3].

In this paper we describe the results of our new, *exact* nonperturbative renormalization group (RG) to study the exotic fractal characteristics of QP systems modelled by a discrete Schrödinger equation. With the exception of some results in § § 4 and 5, most of our results have appeared in our previous publications [4–8]. This review brings out the commonality between a variety of systems and hence puts our results on a more general perspective. Furthermore, it proves the usefulness of our RG methodology in quasiperiodic systems. We hope that this review will serve as a useful source in the literature on incommensurate systems.

We will describe a variety of systems which can be written in the tight binding form

$$a(i)\psi_{i+1} + b(i)\psi_{i-1} + c(i)\psi_i = 0. \tag{1}$$

Here, ψ_i is the fermion wave function and a, b, and c are site-dependent real or complex functions determined by the potential and the eigenvalue. We assume the quasiperiodicity is such that it can be described by a single frequency σ relative to the underlying lattice, i.e., a, b, and c are essentially functions of the fractional part of $i\sigma$ where σ is irrational. Our formulation is valid also for the case where ψ_i is a multi-component vector and a, b, and c are matrices.

Our exact decimation scheme takes into account the inherent quasiperiodicity of the system. The resulting formulation provides a very precise criterion to distinguish the E, C and L phases and hence obtaining the phase diagram. Furthermore, different QP potentials may be described by the same asymptotic RG flow which implies universality for the scaling of the fractal spectra and wave functions. In general, the scaling of the wave functions is determined by the asymptotic RG fixed point or limit cycle whereas the scaling exponents for the spectrum are obtained from the eigenvalues of the linearized RG transformation. The existence of such finite RG attractors provides a new and very accurate method to compute e.g., eigenvalues or phase boundaries. As we will show, in some cases the RG flow is attracted by an infinite RG strange set. Thus, the nonlinearity of the RG transformation opens the way to chaotic dynamics even though the original tight binding model (TBM) is linear.

One of the key questions that we discuss here is the fluctuations of localized wavefunctions. Localization of electrons in disordered materials has been one of the central problems in condensed matter physics. Various recent publications [9] have addressed the problem of scaling of localized wave functions for the case of random disorder and have speculated that exponentially localized wave functions may exhibit fractal fluctuations. However, the issue has remained controversial as these numerical studies have been plagued by finite size effects. We study the fluctuations of the localized wave functions in QP systems using our exact RG approach. The motivation for studying such deterministic disorder is that in the case of random disorder additional complexity arises from distinguishing between the space fluctuations of a single wave function and the disorder fluctuations coming from various different realizations of the randomness. Furthermore, our new RG scheme for QP models can be implemented upto machine precision and is not handicapped by finite size effects encountered in the previous studies.

In this paper, we concentrate on the Harper (10) and its various generalizations which exhibit metallic, insulating, and *critical* phases. Our main focus is the universal scaling behavior in the C phase as well as on the universal self-similar fluctuations of the L phase. We discuss two different generalizations of the Harper equation: one is the TBM associated with two-dimensional electron gas with next nearest neighbor (NNN) interaction and the other is the vector TBM associated with anisotropic quantum XY model in a transverse field. In these generalized models the critical point is replaced by a n a finite parameter interval leading to additional richness and various lasses.

res of the RG method are summarized in § 2. In § 3, we briefly review ion and describe the scaling associated with critical and localized ection 4 deals with the generalized Harper equation in the two- coupling case. The resulting TBM can be either real or complex isotropy properties of the NNN couplings, and we discuss the phase cases. The real TBM exhibits a fat C phase described by a strange G flow. On the other hand, the complex TBM provides an interesting trant metallic phase. In § 5, we discuss the vector TBM for the spinless tation of the anisotropic quantum XY model in an inhomogeneous he exchange anisotropy causes this model to deviate from the Harper tens the critical point into a C phase. We also elaborate on the een the TBM for an electron on the triangular lattice and the quantum transverse field. Finally, in § 6 we summarize our conclusions.

ethod

y developed a new *exact* RG approach which takes into account the y σ of the system. For details of the RG approach, we refer the readers to s [4–8]. Our methodology is somewhat similar to that of Ostlund *et al* es the scaling properties of the wave functions for a specific eigenvalue. thod of Ostlund *et al* which was based on transfer matrices, had limited he fact that at the localization threshold infinite products of transfer d. We propose a scheme where instead of multiplying transfer matrices, is decimated. The main advantage is the reduction in the number of to carry out the renormalization. The price we have to pay is that our s will be slighly more complicated. However, it turns out that with fewer able to eliminate directions which lead to divergences. This not only hing the RG problem but also provides practical means of calculating quantities like the localization threshold and eigenvalues.

n method can be implemented for any irrational σ but here we consider e where σ is given by the inverse golden ratio $\sigma = (\sqrt{5}-1)/2$. In this priate to decimate out all sites except those labelled by the Fibonacci ich are the best rational approximants of the golden ratio). At the nth l, the TBM is expressed in the form

$$\psi(i+F_{n+1}) = \psi(i+F_n) + e_n(i)\psi(i). \tag{2}$$

perty $F_{n+1} = F_n + F_{n-1}$ of the Fibonacci numbers provides exact recur- r the decimation functions e_n and f_n:

$$(i) = -\frac{Ae_n(i)}{1+Af_n(i)}, \tag{3}$$

$$(i) = \frac{f_{n-1}(i+F_n)f_n(i+F_n)}{1+Af_n(i)}, \tag{4}$$

(1) Often the iteration of the recursion relations is numerically more stable than that of ([TBM]).

(2) The possible self-similarity observed by monitoring the behaviour of ψ_i over the range $i \in [-F_n, F_n]$ with increasing n can be captured by a simple asymptotic limit cycle for the decimation functions e_n and f_n. This in turn determines the universal scaling ratios

$$\zeta_j = \lim_{n\to\infty} |\psi(F_{pn+j})/\psi(0)|; \quad j = 0, \ldots, p-1, \tag{5}$$

where p is the cycle length.

(3) An unknown or inaccurate parameter in the TBM and the asymptotic limit cycle can be simultaneously determined self-consistently. This is based on using the secant method to calculate the parameter value for which e.g., $f_n(0) = f_{n-p}(0)$ for some very high n. In other words, the decimation equations themselves provide a new method to compute eigenvalues or phase boundaries up to machine precision.

3. Harper equation

The Harper equation [10], which describes a wide variety of condensed matter problems including the two-dimensional electron gas in a transverse magnetic field, has been in the forefront of theoretical research for the last two decades. In addition to various numerical and RG studies [1], the model has also been solved using the Bethe-ansatz [12].

The Harper equation has the simple form

$$\psi_{i+1} + \psi_{i-1} + 2\lambda\cos[2\pi(i\sigma + \phi)]\psi_i = E\psi_i. \tag{6}$$

The parameter σ describing the magnetic flux in the two-dimensional electron problem introduces quasiperiodicity into the system and is taken as the inverse golden ratio $\sigma = (\sqrt{5}-1)/2$. Although the phase diagram is the same for all diophantine σ, the choice of the golden mean simplifies the RG analysis. Furthermore, it is assumed that the phase ϕ is chosen so that the main peak of the wave function ψ_i is at $i = 0$ [5, 11]. The variation in λ leads to the metal-insulator transition. At the onset of transition $\lambda = 1$, the model has been shown to be self-dual [13] with critical states and a fractal butterfly spectrum [14].

Figure 1 shows the wave function with $E = 0$ at the critical point $\lambda = 1$. The wavefunction repeats itself at every third Fibonacci site. In fact, this repetition of the wave function takes place at all sites: i.e., for every given site, there exists a whole sequence of sites displaced by Q_k (related to the rational approximants of σ^3), where the wave function approaches the same amplitude. [5, 6] This type of translational symmetry in the wave function on a *Fibonacci-lattice* or *Q_k-lattice* is a characteristic feature of the QP systems. Therefore, the wave function does not vanish asymptotically but there are an infinite set of universal numbers ζ (some of them are defined in 5) describing the amplitude of the wave function at points which are spaced further and further apart from

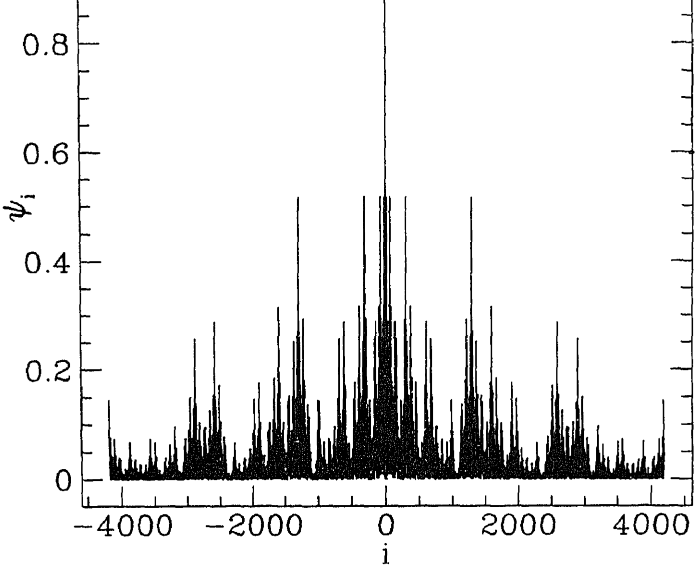


Figure 1. Wave function of the Harper equation at the critical point $\lambda = 1$ with $E = 0$ and $\phi = 1/4$.

the central peak. In addition, the structure around the subpeaks mimics the behavior seen on a larger length scale resulting in self-similarity in the wave function.

In the RG approach, the self-similarity results in a nontrivial asymptotic p-cycle for the decimation functions and the scaling ratio. In the E phase, the RG cycle is trivial with $\zeta = 1$. It should be noted that eigenstates with different energies are characterized by different asymptotic behavior of the RG trajectories. In addition to $E = 0$, the eigenstates at the band edges are described by a limit cycle while almost all eigen energies are believed to correspond to a strange invariant set.

In order to investigate the localized states, the wave function ψ_i is written as

$$\psi_i = \mathrm{e}^{-\gamma|i|}\eta_i, \tag{7}$$

where γ is the Lyapunov exponent which vanishes in the E and C phase. The L phase is characterized by a positive Lyapunov exponent $\gamma = \ln(\lambda)$ corresponding to an exponential decay of the wave function [AA]. Therefore, the function η_i describing the fluctuations of the localized wave function satisfies the following TBM for $i > 0$:

$$\frac{1}{\lambda}\eta_{i+1} + \lambda\eta_{i-1} + 2\lambda\cos[2\pi(i\sigma + \phi)]\eta_i = E\eta_i. \tag{8}$$

We study this TBM using the RG method. Figure 2 shows the fluctuations η_i at a band maximum (the band center with $E = 0$ is described by a period-3 RG cycle). The fluctuations for $\lambda > 1$ possess the same complexity and richness as the critical states. Furthermore, it turns out that the self-similar fluctuations are universal throughout the L

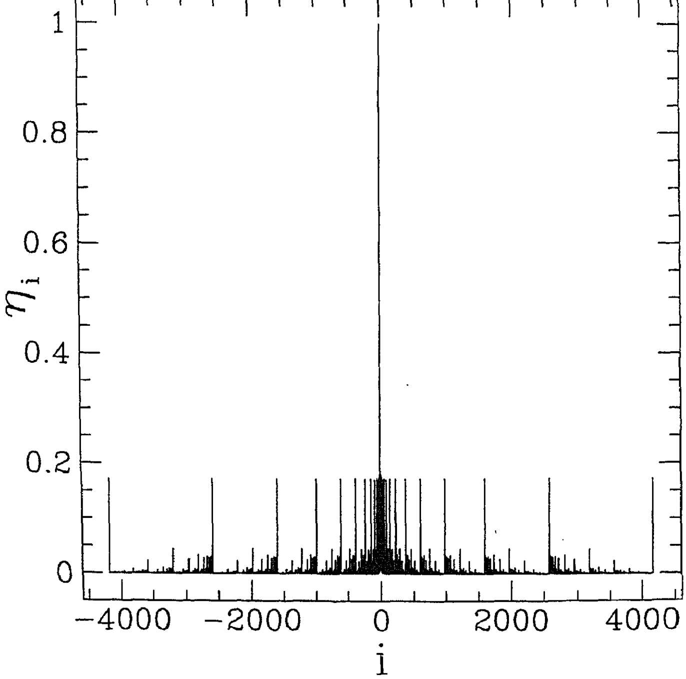


Figure 2. Absolute value of self-similar universal fluctuations in the supercritical phase of the Harper equation.

phase and are characterized by a unique strong coupling RG fixed point with $\zeta = 0.172586\cdots$. In particular, after writing the RG transformation and the decimation functions in the form where the discrete lattice index i is replaced by a continuous renormalized variable, we are able to solve for a power series expansion of the fixed point by the Newton method. Details of the RG fixed point and the eigenvalue analysis are discussed elsewhere. [8] It should be noted that although the fluctuations bear a qualitative resemblance to the wave functions in the C phase, quantitatively they belong to different universality classes.

4. Generalized Harper equation: Bloch electrons with NNN interaction

The Harper equation models a two-dimensional Bloch electron on a square lattice in a transverse magnetic field when only the hopping to nearest-neighbor (NN) lattice sites is taken into account. Including the hopping also to the NNN sites one obtains the following TBM:

$$\begin{aligned} &t_a(\psi_{k+1}+\psi_{k-1}) + 2t_b\cos[2\pi(k\sigma+\phi)]\psi_k \\ &\quad + \exp[i2\pi(k\sigma+\phi)]\{t_{ab}e^{i\pi\sigma}\psi_{k+1} + t_{a\bar{b}}e^{-i\pi\sigma}\psi_{k-1}\} \\ &\quad + \exp[-i2\pi(k\sigma+\phi)]\{t_{ab}e^{i\pi\sigma}\psi_{k-1} + t_{a\bar{b}}e^{-i\pi\sigma}\psi_{k+1}\} = E\psi_k\,. \end{aligned} \tag{9}$$

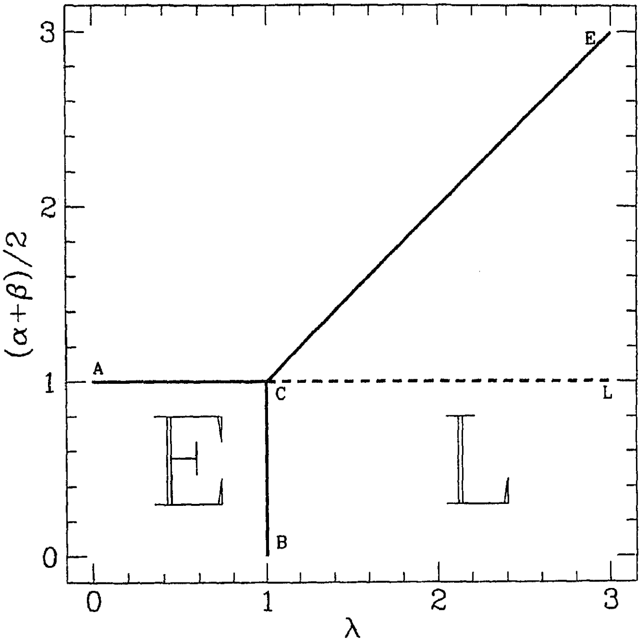


Figure 3. Phase diagram of the two-dimensional electron gas with NNN interaction. For $t_{ab} = t_{a\bar{b}}$, the region bounded by the AL and CE lines is critical. However, as soon as these two NNN couplings become different from each other, the C phase is replaced by a reentrant E phase. The solid lines BC, AC, and CE always remain critical. The extension of the line AC (line CL) separates the L phase into two different regions with different scaling properties.

Here t_a and t_b are the NN couplings while t_{ab} and $t_{a\bar{b}}$ are the diagonal NNN couplings which are taken to be zero in the case of the Harper equation. In the isotropic limit $t_{ab} = t_{a\bar{b}}$, the TBM becomes real and has been studied in various recent papers [15, 7, 8]. The two cases $t_{ab} = t_{a\bar{b}}$ and $t_{ab} \neq t_{a\bar{b}}$ will be discussed separately below.

Case I. $t_{ab} = t_{a\bar{b}}$. In this limit, the above TBM can be written as

$$\begin{aligned}&\{1+\alpha\cos[2\pi(\sigma(i+\tfrac{1}{2})+\phi)]\}\psi_{i+1}+\{1+\alpha\cos[2\pi(\sigma(i-\tfrac{1}{2})+\phi)]\}\psi_{i-1}\\&\quad+2\lambda\cos[2\pi(\sigma i+\phi)]\psi_i = E\psi_i\,. \qquad (10)\end{aligned}$$

Here $\lambda = t_b/t_a$ and $\alpha = 2t_{ab}/t_a$. The Harper equation corresponds to the limit $\alpha = 0$ where the NNN coupling term is zero.

Figure 3 shows the phase diagram of the model in the $\lambda - \alpha$ plane [15, 7]. By carrying out the RG analysis at the band minimum, we find that the model belongs to the universality class of the Harper equation for $\alpha < 1$ [7]. Therefore, along the line BC (except at the point C), the universal scaling is identical to that of the Harper equation. An interesting feature of the phase diagram is that the C phase exists in a finite region of parameter values. Unlike the scaling at the critical point of the Harper equation given by limit cycles of the RG equation, the scaling in this

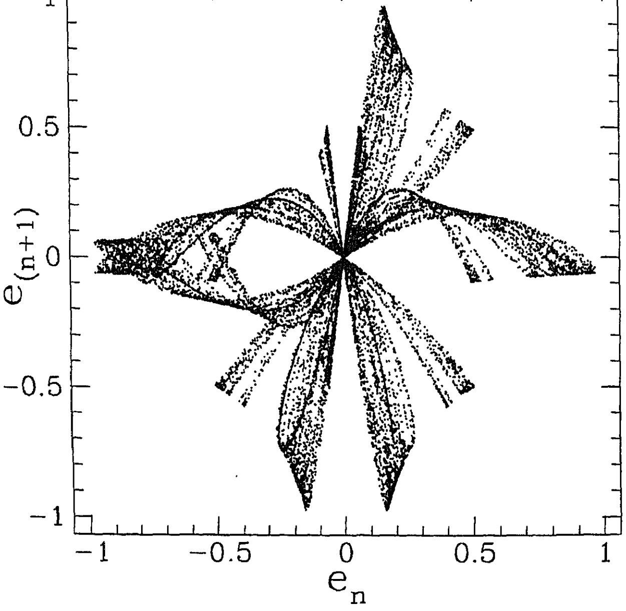


Figure 4. A two-dimensional projection of the attractor of the renormalization flow in the strong coupling limit $\lambda \to \infty$ with $\alpha = \beta > \lambda$. For about 4000 different parameter values, decimation equations were iterated 35 times and the first 10 iterates were ignored as transients.

C phase was found to be characterized by an infinite strange set of the RG flow [7]. Therefore, the fractal wave functions do not exhibit self-similarity and are described by an ergodic set which brings out the order underlying the chaotic RG trajectories.

It turns out that the qualitative behavior of the RG trajectories describing the fluctuations in the L phase are similar to that of the C phase discussed above. The universality class for the fluctuations in the L phase turns out to be unaltered as long as $\alpha < 1$ [8]. Above the threshold $\alpha = 1$ (above the line CL), the situation is again much more complicated. Almost all parameter values in this region are described by the same infinite ergodic set of scaling ratios. In analogy with the C phase discussed above, the ergodic set of scalings, when projected on two dimensions, exhibit an intriguing order manifested in the form of an orchid-flower (see figure 4).

Another intriguing feature of the C and L phase is that there appears to be a special set of parameter values $\alpha = 1/|\cos(2\pi r)|$ with rational r where the RG trajectories in both the L and C phase converge on a limit cycle of a high period (for a fixed α, the same limit cycle is observed for all λ in the L phase whereas in the C phase also λ has to be chosen carefully to obtain a finite cycle). We conjecture that these unstable fixed points are dense on the strange attractor of the RG flow.

Detailed numerical studies of RG flow show that the lines in the phase diagram describing the boundaries between the E, C and L phases exhibit fractal wave functions which differ quantitatively from the wave functions in the C phase discussed above. The lines AC (E-C transition) and CE (C-L transition) appear to correspond to their own RG strange sets [7]. The attractor associated with the line CE seems to envolope the strange set of the C phase enclosed by AC and CE.

Case II. $t_{ab} \neq t_{a\bar{b}}$. In this case, the TBM is complex and has not been fully investigated previously. Using the duality property of the model, Han *et al* [15] calculated the Lyapunov exponent of the model analytically in the space of the three parameters $\lambda = t_b/t_a$, $\alpha = 2t_{ab}/t_a$, and $\beta = 2t_{a\bar{b}}/t_a$. They concluded that the system is localized for $\lambda > 1$ if $(\alpha+\beta)/2 < 1$ and for $\lambda > (\alpha+\beta)/2$ otherwise. Apart from the existence of the metal-insulator transition, nothing was known about the scaling properties of the complex model. In particular, one of the important questions is how the inequality in the NNN couplings affects the scaling properties in the C and L phases described by the strange invariant sets.

With the decimation methodology, we found that the phase diagram changed discontinuosly as $\alpha - \beta$ became different from zero. As soon as α and β differ, $\alpha - \beta$ was found to be an irrelevant parameter and the phase diagram is determined solely by two parameters: λ and $(\alpha+\beta)/2$.

In the parameter range $(\alpha+\beta)/2 < 1$, the decimation functions become asymptotically real approaching the same universal cycles as for the Harper equation $\alpha = \beta = 0$. For $\alpha + \beta \geq 1$, $\alpha \neq \beta$, the decimation functions stay complex also asymptotically. Here the C phase of the case $\alpha = \beta > 1$ is replaced by a reentrant E phase. Moreover, the E-L transition belongs to the universality class of the corresponding transition in the Harper equation upto a complex phase factor in ζ. The invariant strange sets of the renormalization describing the fluctuations in the L phase for the limit $\alpha = \beta$ degenerates into the strong coupling Harper fixed point as soon as α and β become different [16].

On the line $(\alpha+\beta)/2 = 1$, the model can be explicitly shown to be related to the TBM for the triangular lattice, upto a complex phase factor [16]. The TBM describing Bloch electrons on a triangular lattice is [17]

$$\cos\left[\pi\left(i\sigma + \left(\phi + \frac{\sigma}{2}\right)\right)\right]\psi_{i+1} + \cos\left[\pi\left(i\sigma + \left(\phi - \frac{\sigma}{2}\right)\right)\right]\psi_{i-1} + \lambda\cos[2\pi(i\sigma + \phi)]\psi_i = E\psi_i. \tag{11}$$

Our detailed RG analysis shows that unlike the E-L transition of the TBM describing the square lattice, the TBM describing the triangular lattice exhibits the C-L transition: i.e., the line AC remains critical (described by a unique fixed point except at the point C which is characterized by its own fixed point) while the wave functions along the line CL are localized. The fractal characteristics on the line AC, point C and the fluctuations on the line CL define new universality classes (of the triangular lattice) which are quantitatively different from those of the square lattice, namely the Harper universality classes. The reentrant E phase can be understood in the anisotropic limit, $\alpha \to \infty$, $\beta/\alpha \to 0$, and $\lambda/\alpha \to 0$.

The resulting TBM at a site k is related to the weak coupling limit of the Harper model by a complex phase factor $\exp(i2\pi\phi k)\exp(i\pi\sigma k^2)$ [16].

We next describe a related problem, namely the XY quantum spin chain, described by a vector TBM which can be viewed as a perturbed Harper equation

$$J(y,x)\Psi_{i+1} + J(x,y)\Psi_{i-1} + 2h_i\Psi_i = \bar{E}\Psi_i, \tag{12}$$

where $h_i = \lambda\cos(2\pi(i\sigma + \phi))$, $J(x,y)$ and $\bar{E}$ are 2×2 matrices and Ψ is a two-dimensional vector describing the coupled system

$$J_y\psi_{i+1} + J_x\psi_{i-1} + 2h_i\psi_i = E\eta_i, \tag{13}$$

$$J_x\eta_{i+1} + J_y\eta_{i-1} + 2h_i\eta_i = E\psi_i. \tag{14}$$

The above model is the fermion representation of the quantum XY spin chain in a transverse field with the Hamiltonian [18]

$$H = -\sum_i \left[\tfrac{1}{2}(J_x s_i^x s_{i+1}^x + J_y s_i^y s_{i+1}^y) + h_i s_i^z\right]. \tag{15}$$

Here s are the Pauli matrices describing the spin 1/2. J_x and J_y describe ferromagnetic exchange interactions. The model can also be written as

$$(H_h - 2gH_g)(H_h + 2gH_g)\Psi = E^2\Psi, \tag{16}$$

where H_h denotes the Harper Hamiltonian and H_g denotes the part of the Hamiltonian which exists only at a non-zero value of the anisotropy g defined as $J_x = 1 - g$ and $J_y = 1 + g$. We see that the anisotropy perturbs the "squared" Harper model.

Figure 5 shows the phase diagram in the $\lambda - g$ plane [19, 20]. Whenever λ and one of the exchange interactions become equal either the E-C or C-L transition takes place. The E-C transition corresponds always to a smaller absolute value of λ, i.e. to the exchange interaction whose absolute value is smaller. Whenever J_x and J_y differ a fat C phase is observed in the phase diagram. Therefore, the fattening of the critical point is due to the breaking of the $U(1)$ symmetry in the fermion Hamiltonian which is a consequence of the broken $O(2)$ symmetry in spin space.

An interesting consequence of the spectral and magnetic interplay is the fact that the onset of the C-L transition is coincident with the onset of the magnetic transition to the long range order (LRO) where two-point long range spin-spin correlations vanish. However, the significance of the E-C transition on the magnetic properties of the quantum chain remains a mystery.

The C phase of the model is found to be described by three different universality classes [6]. The E-C transition is characterized by a different universality class from that of the C-L transition line. The critical phase sandwiched between the E-C and C-L transitions is conjectured to form its own universality class although the existence of a RG cycle has been confirmed only in the Ising limit $J_y = 0$ (see the next section). All these three universality classes are different from that of the Harper critical point.

The fluctuations in the L phase are believed to be universally characterized by a strong coupling fixed point. This was confirmed in the Ising limit discussed in the next section.

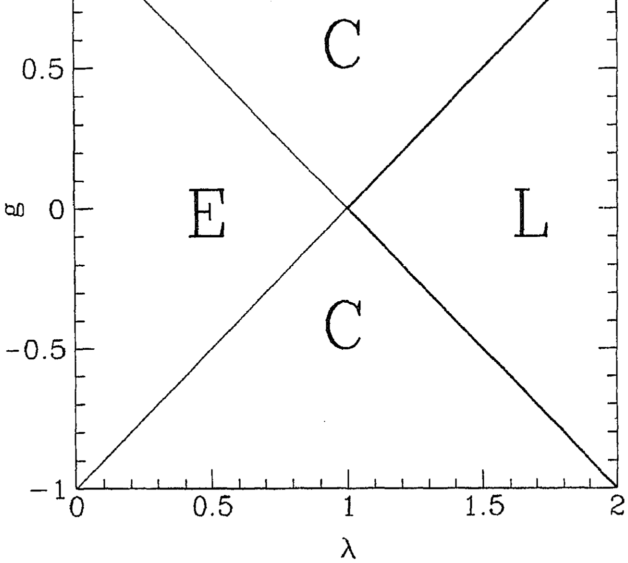


Figure 5. Phase diagram for the quasi-particle excitations of the quantum spin chain with $J_x = 1 - g$ and $J_y = 1 + g$. For $g = \pm 1$, the model reduces to Ising model with C and L phases. The dark boundary line separating C and L phases is the conformal line describing the onset to magnetic long range order.

5.2 *Relation to the bloch electrons*

It is rather interesting that the Bloch electron problem and the spin problem are related in certain limits. The TBM for the triangular lattice (see (11)) resembles the TBM for the Ising limit ($J_y = 0$) of the quantum spin chain discussed in the previous section [19, 20]. Both are modeled by a NN TBM where the diagonal as well as the off-diagonal terms are modulating. The periodicity of the diagonal term is twice the periodicity of the off-diagonal term. The only difference between these two models is in the relative phase differences between the diagonal and off-diagonal terms. As a consequence, the wave functions at the band edges have exact symmetry in the triangular model but only asymptotic symmetry in the quantum Ising model (QIM).

The phase diagram is the same for both systems with C phase for $\lambda \leq 1$ and the L phase for $\lambda > 1$ (see figure 5 at $g = -1$). There are three RG limit cycles which determine the asymptotic scaling of the wave functions at the upper or lower band edges in both systems for $\lambda < 1$, $\lambda > 1$, and $\lambda = 1$, respectively. However, an exception to this rule is formed by the lower band edge of the QIM at $\lambda = 1$ which is described by its own RG fixed point. The corresponding state is believed to be conformally invariant mapping to none of the quantum states of the triangular model [5, 6].

The fact that RG analysis singles out the conformal state in distinguishing the two models that bear striking mathematical resemblance (inspite of some important physical of the physics underlying the QP models discussed here.

Another interesting feature is that the conformally invariant state of the Ising model also describes the fluctuations in the exponentially localized strong coupling limit of the Harper equation for $E = 0$ [8]. In other words, the self-similar fluctuations at the band center of the supercritical Harper model describe the self-similarity of the quantum state responsible for the long range order in Ising spin. This fact can be understood by noticing that at the conformal point, the TBM described by (13) resembles equation (8) describing the fluctuations in the localized phase of the Harper equation. Therefore, the excitations in the spin model are related to the fluctuations in the localized states of the Harper equation and the conformal universality class of the spin model is related to the strong coupling universality of the Harper equation.

6. Conclusions

Our study of a variety of systems with a common feature being the competing length scales, shows the existence of interesting phase diagrams exhibiting E-L, E-C and C-L transitions. Furthermore, another intriguing possibility is the reentrant E phase. This richness in the phase diagram manifests itself in interesting variations in the transport properties as the parameters of the system are varied. We believe that our results can be experimentally realized on two-dimensional mesoscopic systems. The results on the quantum spin chain may be relevant in magnetic superlattices.

Wave functions in both the *critical* and localized phases exhibit fractal character. The universal scaling properties in these phases can be studied with very high precision using a decimation method. The decimation scheme unveils some interesting relationships between seemingly unrelated problems and limits the number of universality classes. One of our important results is that a wide class of systems can be shown to belong to the universality class of the NN square lattice (Harper universality) or the triangular lattice. In addition to determining the phase diagram and scaling characteristics, our method provides a new tool to determine eigenvalues, phase boundaries, and Lyapunov exponents. This is because unknown parameters in the TBM can be determined self-consistently with the RG limit cycle upto machine precision. We refer the readers to our earlier papers for various detailed examples.

Another important result of our studies that the L phase exhibit the same richness as the C phase with fractal characteristics, has to be viewed from a more general perspective. The localization phenomenon in QP systems is analogous to the Anderson localization of the random systems. Therefore, our findings strengthen the likelyhood of fractal characteristics in other random or aperiodic systems exhibiting localization.

Finally, we would like to point out that there are many other systems in both classical and quantum chaos [24] which can be modeled by [TBM]. One interesting example is the study of phonon modes in the classical Frenkel-Kontorova model exhibiting the pinning-depinning transition [22, 23]. Our very recent studies have shown the existence of a new type of phonon modes with eigenstates represented by an infinite series of step functions [24]. Another important problem that is under investigation is the quantum kicked rotor which has been mapped to a TBM [21]. Thus, we believe that there are many more interesting results to be discovered using the RG methods.

Acknowledgements

The research of IIS is supported by a grant from National Science Foundation DMR 093296. IIS would like to acknowledge the hospitality of National Institue of Standard and Technology where part of this work was done. JAK would like to thank for the support by the Niilo Helander Foundation and for the hospitality during his visits to the George Mason University where this work was initiated.

References

[1] For a review, see J B Sokoloff, *Phys. Rep.* **126**, 189 (1985)
[2] V I Arnold and A Avez, *Ergodic problems of classical mechanics* (Benzamin, New York, 1968)
[3] B Simon, *Adv. Appl. Math.* **3**, 463 (1982)
[4] J A Ketoja, *Phys. Rev. Lett.* **69**, 2180 (1992)
[5] J A Ketoja and I I Satija, *Phys. Lett.* **A194**, 64 (1994)
[6] J A Ketoja and I I Satija, *Physica* **A219**, 212 (1995)
[7] J A Ketoja, I I Satija and J C Chaves, *Phys. Rev.* **B52**, 3026 (1995)
[8] J A Ketoja and I I Satija, *Phys. Rev. Lett.* **75**, 2762 (1995)
[9] L Pietronero, A P Siebesma, E Tosatti and M Zannetti, *Phys. Rev.* **B36**, 5635 (1987)
M Schreiber and H Grussbach, *Phys. Rev. Lett.* **67**, 607 (1991)
F Lema and C Wiecko, *Phys. Scr.* **47**, 129 (1993)
[10] P G Harper, *Proc. Phys. Soc. London* **A68**, 874 (1955)
[11] S Ostlund and R Pandit, *Phys. Rev.* **B29**, 1394 (1984)
S Ostlund, R Pandit, D Rand, H J Schellnhuber and E D Siggia, *Phys. Rev. Lett.* **50**, 1873 (1983)
[12] P B Wiegmann and A V Zabrodin, *Phys. Rev. Lett.* **72**, 1890 (1994)
L D Faddeev and R M Kashaev, *Commun. Math. Phys.* **169**, 181 (1995)
[13] S Aubry and G André, *Ann Israel Phys. Soc.* **3**, 133 (1980)
[14] D R Hofstadter, *Phys. Rev.* **B14**, 2239 (1976)
[15] J H Han, D J Thouless, H Hiramoto and M Kohmoto, *Phys. Rev.* **B50**, 11365 (1994)
[16] J A Ketoja and I I Satija, *Reentrant phase diagram of the generalized Harper equation*, preprint (1996); *J. Phys.* **C9**, 1123 (1997)
[17] F H Claro and G H Wannier, *Phys. Rev.* **B19**, 6068 (1979)
D J Thouless, *Phys. Rev.* **B28**, 4272 (1983)
[18] E Lieb, T Schultz and D Mattis, *Ann. Phys. (N.Y.)* **16**, 407 (1961)
[19] I I Satija *Phys. Rev.* **B48**, 3511 (1993); *Phys. Rev.* **B49**, 3391 (1994)
I I Satija and M Doria, *Phys. Rev.* **B39**, 9757 (1989)
[20] I I Satija and J C Chaves, *Phys. Rev.* **B49**, 13239 (1994)
[21] S Fishman, D R Grempel and R E Prange, *Phys. Rev. Lett.* **49**, 509 (1982)
[22] P Bak, *Rep. Prog. Phys.* **45**, 587 (1982)
[23] S Aubry, *Physica* **D7**, 240 (1983)
[24] J A Ketoja and I I Satija, *Critical phonons of the supercritical Frenkel–Kontorova model: Renormalization bifurcation diagrams*, *Physica* **D** (in press)

Dynamical signatures of 'phase transitions': Chaos in finite clusters

VISHAL MEHRA, SAROJ K NAYAK* and RAMAKRISHNA RAMASWAMY
School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067, India
*Department of Physics, Virginia Commonwealth University, Richmond, VA 23284, USA

Abstract. Finite clusters of atoms or molecules, typically composed of about 50 particles (and often as few as 13 or even less) have proved to be useful prototypes of systems undergoing phase transitions. Analogues of the solid–liquid melting transition, surface melting, structural phase transitions and the glass transition have been observed in cluster systems. The methods of nonlinear dynamics can be applied to systems of this size, and these have helped elucidate the nature of the microscopic dynamics, which, as a function of internal energy (or 'temperature') can be in a solidlike, liquidlike, or even gaseous state. The Lyapunov exponents show a characteristic behaviour as a function of energy, and provide a reliable signature of the solid–liquid melting phase transition. The behaviour of such indices at other phase transitions has only partially been explored. These and related applications are reviewed in the present article.

Keywords. Clusters; chaos; phase transitions; Lyapunov exponents.

PACS Nos 5.45; 64.70; 36.40

1. Introduction

One major impetus for research in the dynamics of nonintegrable systems has been the hope of understanding the basis for the effectiveness of statistical mechanical ideas in the description of systems with several degrees of freedom [1]. On the one hand, while the ergodic hypothesis is presumed to be valid for generic many-body Hamiltonians, it is difficult to prove for any particular system since a realistic dynamical picture of a multiparticle system is not easy to obtain owing to the extremely large number of degrees of freedom. On the other hand, the application of ideas of statistical mechanics to few-body systems is fraught with difficulty.

A regime where some progress can be made has been provided in recent years by cluster systems, namely aggregates of atoms or molecules consisting of typically between 10 and 1000 particles [2]. These systems are small enough that the explicit dynamics can be followed for reasonably long times. They are also large enough for it to be possible to characterise them as having a well-defined temperature or being in a particular state of matter. Indeed, in recent years, a major interest in cluster studies is the elucidation of simple phase transformations [3, 4]. The study of the dynamics at these phase transformations is the subject of this article.

Clusters, being intermediate in size between molecules and bulk matter, have physical and chemical properties which can sometimes be very different from the properties of a

Theoretical studies of the dynamics of cluster systems are usually carried out using classical molecular dynamics (CMD) techniques. The connection between the microscopic dynamics and thermodynamics or statistical mechanics can be explored in some detail by studying clusters of increasing size or varying other parameters of the system such as particle mass or the interaction potential. These studies also play an important role in clarifying the spread of ergodic behavior in many-body systems.

In the context of nonlinear studies, finite clusters are important as they provide useful examples of micro- and mesoscopic systems where the methods of nonlinear dynamics (which were essentially developed with low-dimensional systems in mind) can find application. For a N-particle cluster, the number of freedoms is $3N$, which is quite large for even the smallest ($N \approx 7$) systems studied in this context. Given the highly nonlinear nature of the interparticle interaction, the motion is unlikely to be globally regular at finite temperatures. Thus, chaotic dynamics may provide the basic mechanism for inducing statistical behavior in cluster systems by facilitating the rapid mixing among modes and consequent redistribution of vibrational energy. In general, stochastic trajectories are presumed to be the main cause for giving rise to statistical behavior, when the ergodic hypothesis, namely that the time averages are identical to the phase space averages, is expected to hold. This statistical description is manifest in the measured values and the distributions of different quantities [9–11].

With regard to simple phase transitions, it is now well-established that although the numbers of atoms or molecules present is small compared to bulk material ($\sim 10^{23}$), clusters can exhibit structural isomerization [12–15] and related phenomena. Small clusters go from a solid-like phase to a liquid-like phase as the temperature or total internal energy is increased [16]. In bulk matter, the temperature at which the solid melts, T_{m}, is same as the temperature at which the corresponding liquid freezes, T_{f}. However, T_{m} and T_{f} can be different in finite clusters [17], when there can be dynamical coexistence [18, 19] of both of these phases in the intermediate temperature regime (between T_{f} and T_{m}) when clusters start melting. This phenomenon is distinct from the phase change in bulk matter, and has been experimentally verified; extensive molecular dynamics and Monte Carlo simulations [16] also support this picture. Cluster systems can show yet another form of coexistence in surface melting where atoms in the cluster core are at a lower effective temperature and are in a solid state, while atoms on the surface are in a liquid-like state and show much greater diffusivity [13, 20]. At temperatures below the melting transitions, there can also be structural phase transitions [21]. In addition, the so-called glass transition can also be mimicked by cluster systems [22].

The methods and techniques of nonlinear dynamics of relevance to the study of atomic clusters are reviewed in this article. These include the computation of Lyapunov exponents, in particular the largest (maximal) one which has been found to be a reliable phase-space indicator of cluster phase change, the Kolmogorov (K-) entropy, short time distributions of local K-entropy and Lyapunov exponents and Lyapunov spectra. These prove invaluable in understanding phase transitions since there are few available reliable measures for exploring the process of phase change in finite systems.

In the next section we briefly describe the classical methodology of cluster simulations, and define the Lyapunov exponents and the K-entropy. We then discuss the connection between phase transformations (solid-to-liquid) and the maximal Lyapunov exponent (MLE). This is followed by a review of some of the other techniques that have been applied in cluster dynamical studies. We conclude in § 5 with a summary.

2. Cluster dynamics, Lyapunov exponents and the Kolmogorov entropy

The simulation of clusters is usually performed through Monte Carlo (MC) methods, or through molecular dynamics (MD) [23], namely the integration of the relevant Hamilton's or Newton's equations of motion. MD simulations can be carried out either at constant energy, when the microcanonical ensemble obtains, or at constant temperature, with the use of thermostats, in which case the canonical ensemble results. For dynamical studies the most relevant are the MD methods although some progress is being made in the extraction of dynamical information from equilibrium MC simulations [24].

Details of the molecular dynamics simulations of N-particle systems have been extensively documented [23], and the computational methodology of integrating Newton's equation of motion has been developed in considerable detail. Since a cluster is an isolated group of particles, the boundary conditions are simpler to incorporate than in bulk or liquid-state simulations. The interaction between particles is normally taken to be composed of pair-potentials, so that the total potential energy $\mathcal{V}$ for a configuration $\mathbf{R} \equiv r_1, r_2, \ldots, r_N$ is

$$\mathcal{V}(\mathcal{R}) = \sum_{i<j}^{N} V(r_{ij}) \tag{1}$$

where r_i is the position of the i th particle and $r_{ij} \equiv |r_i - r_j|$ is the distance between particles i and j. The pair-potential $V(r)$ is usually taken to be of a simple analytic form, such as the Lennard-Jones (for rare-gas systems), Yukawa (for nuclear systems) or Gupta (for metallic clusters) [25–28]. The total potential energy, $\mathcal{V}$, is an extremely complicated function of the configuration $\mathbf{R}$. Finding the ground state, namely the global minimum, is a tricky problem, although for small enough clusters, this and the lowest few local minima (corresponding to isomers) can be located by several techniques.

The dynamics follows Newton's equations, which are (m_i is the mass of the ith particle)

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial \mathcal{V}}{\partial \mathbf{r}_i} \quad i = 1, 2, \ldots, N \tag{2}$$

The total energy, namely the Hamiltonian $\mathcal{H} = \mathcal{T} + \mathcal{V}$, where $\mathcal{T} = \sum \frac{1}{2} m_i \dot{\mathbf{r}}_i{}^2$, is conserved, as are all the components of the total momentum and the angular momentum, giving a total of 7 constants of the motion. At any given total energy, the 'temperature' of the cluster is defined through the average kinetic energy as

$$T = \frac{2\langle \mathcal{T} \rangle}{(3N - 6)k_b} \tag{3}$$

k_b being the Boltzmann constant and $\langle\ \rangle$ denoting an average over the entire trajectory.

The spectrum of Lyapunov exponents of a given dynamical system characterizes the dynamical instability of the orbits. An N-freedom hamiltonian system has $2N$ Lyapunov exponents, $\lambda_1, \ldots, \lambda_{2N}$ which are defined by the eigenvalue spectrum of the squared Jacobi matrix [29]

$$J(t) = \frac{\partial(\mathrm{p}(t), \mathrm{r}(t))}{\partial(\mathrm{p}(0), \mathrm{r}(0))}, \qquad (\mathrm{p}_i \equiv m\dot{\mathrm{r}}_i \quad \text{and} \quad \mathrm{p} \equiv \mathrm{p}_1 \cdots \mathrm{p}_N) \tag{4}$$

in the asymptotic limit i.e.,

$$\{\mathrm{e}^{2\lambda_1\tau}, \ldots, \mathrm{e}^{2\lambda_{2N}\tau}\} = \text{eigenvalue spectrum of } J^T(\tau)J(\tau) \text{ as } \tau \rightarrow \infty \tag{5}$$

Therefore, for a N-particle system in three dimensions there are $6N$ Lyapunov exponents, which come in matching positive-negative pairs as a consequence of Liouville's theorem [30]. They are independent of initial condition $(\mathrm{r}(0), \mathrm{p}(0))$ for ergodic Hamiltonians, and further, if there are L independent constants of motion then $2L$ of the exponents vanish. Hence for a cluster system of N particles there will be $3N - 7$ positive Lyapunov exponents. From Pesin's theorem [31] the sum of these equals the Kolmogorov entropy, which measures the rate of loss of information in a chaotic system. Standard techniques to compute the entire spectrum of Lyapunov exponents have been devised, and these can be further simplified to get only the largest one, the MLE [32, 33]. If the distance between two trajectories at time $t = 0$ is $\mathrm{d}(0)$, and the distance at time t is $\mathrm{d}(t)$, then

$$\lambda_1 = \lim_{t\rightarrow\infty} \lim_{d(0)\rightarrow 0} \frac{1}{t} \log \frac{\mathrm{d}(t)}{\mathrm{d}(0)} \tag{6}$$

The actual computation of the MLE uses the above definition, but calculates the rate of increase of distance in tangent space [32].

3. Phase transitions and Lyapunov exponents

The behaviour of the MLE at a phase transition was first studied by Butera and Caravati [34] who simulated the planar O(2) Heisenberg model of interacting planar rotors on a lattice in $2d$ as a function of temperature. They noted that the MLE had a distinct bend or 'knee' at the temperature of the Kosterlitz-Thouless transition.

In subsequent studies, Posch and Hoover [35] studied Lyapunov spectra for monoatomic system in two and three spatial dimensions, both solid and liquid states using a repulisive potential. Their results show that the partial spectrum of positive Lyapunov exponents, λ_n, are well approximated by a power law $\lambda_n = \alpha n^\beta$ with $\beta \approx \frac{1}{3}$ for solid-like states and $\beta \approx 1$ for the liquidlike states. Hinde *et al* [36] examined the entire Lyapunov exponent spectra in 7 atom rare-gas cluster and observed the linear dependence of λ_n with distinct slopes in solidlike and liquidlike states.

The idea that the MLE could be an indicator of phase changes was further investigated by Nayak *et al* [37], who examined the melting of rare-gas clusters by using classical molecular dynamics techniques. Simulations of atomic clusters, using the Lennard-Jones pair potential for cluster sizes of 7, 13 and 55 showed that the MLE has a characteristic dependence at the temperature of the solid-liquid melting transition.

Studies of finite clusters have explored this transition for several years, and the usual manner in which melting is detected is by computing the Lindemann index, namely the average rms fluctuation of the atoms about their equilibrium positions,

$$\delta = \frac{2}{N(N-1)} \sum_{i<j} \frac{\sqrt{\langle r_{ij}^2 \rangle - \langle r_{ij} \rangle^2}}{r_{ij}}, \tag{7}$$

i indexing the various atoms, r_{ij} the interparticle distances and $\langle\ \rangle$ implying an average over a MD trajectory. The Lindemann criterion [38, 39] for bulk melting is that δ is larger than 0.1, and has its origin in empirical observations on the characteristic distances between atoms, as estimated from the amplitude of vibrations, which is greater in the liquid state as compared to the solid. This criterion is used even for the study of small systems and is known to correlate well with other indicators of melting such as diffusivity.

A striking result of the dynamical studies [37] was that the change in the largest Lyapunov exponent occurs at the same total energy when $\delta \approx 0.1$, namely, when the Lindemann criterion for melting was satisfied. At melting, the cluster goes from fairly rigid to nonrigid configurations, and is hence able to access a larger volume of the phase space. As a result, larger nonlinearities become important and the motion becomes globally chaotic. Thus the MLE, which is either zero (or small) for energies below the transition temperature, increases significantly at the transition (figure 1). For clusters such as Ar_{13}, the change in the MLE at the melting transition is large, whereas in smaller clusters such as Ar_7, there is only a discontinuity in the slope as the energy increases. In even smaller clusters such as Ar_3, the MLE increases uniformly with temperature, and there is no explicit signature of a phase transition (Ar_7 is the smallest cluster that can be considered to show any evidence of a change in "phase" with temperature [40]). Comparison with previously published results [18, 41, 42] for various cluster sizes shows that the behaviour of the MLE with energy or temperature closely parallels the behaviour of the Lindemann index.

The magnitude of the MLE may be correlated with the volume of phase space available to the system, which also increases significantly at a phase transition. Such increase of phase-space volume apparently accelerates the divergence of nearby trajectories, and this appears to be a generic feature of nonlinear systems. Since the sum of positive Lyapunov exponents is the Kolmogorov entropy, it is not surprising that the MLE should reflect the large increase in entropy at a phase transition. This connection has been explored to some extent–see e.g. refs. [34, 40, 43]. One also therefore expects that the energy density of states should increase sharply at approximately the energy at which the isoergic MD simulations show a sharp rise in the Lyapunov exponent. This has indeed been noted earlier [44], as an increase in the configurational density of states in the cluster solid-liquid coexistence regime, which in turn is reflected in a bimodal character of PE distributions obtained from canonical Monte Carlo simulations for the phase coexistence temperatures. (The density of states can be constructed from constant temperature canonical ensemble simulations using multiple histogram method; see [37] for details.) Although the discussion above is in terms of the internal energy of the clusters, similar

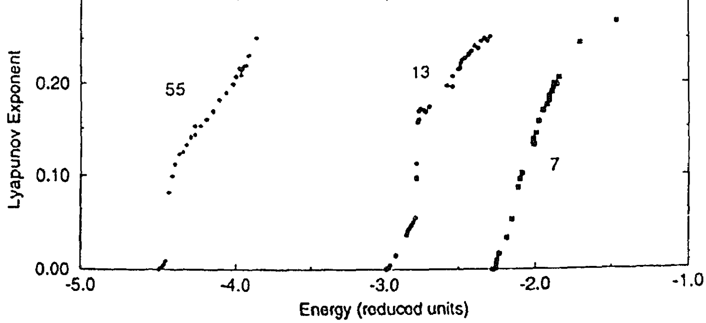


Figure 1. Variation of the largest Lyapunov exponent, λ_1, with internal energy (units are ϵ per atom) for Lennard–Jones clusters of various sizes. The dependence of MLE on the temperature also shows similar behavior.

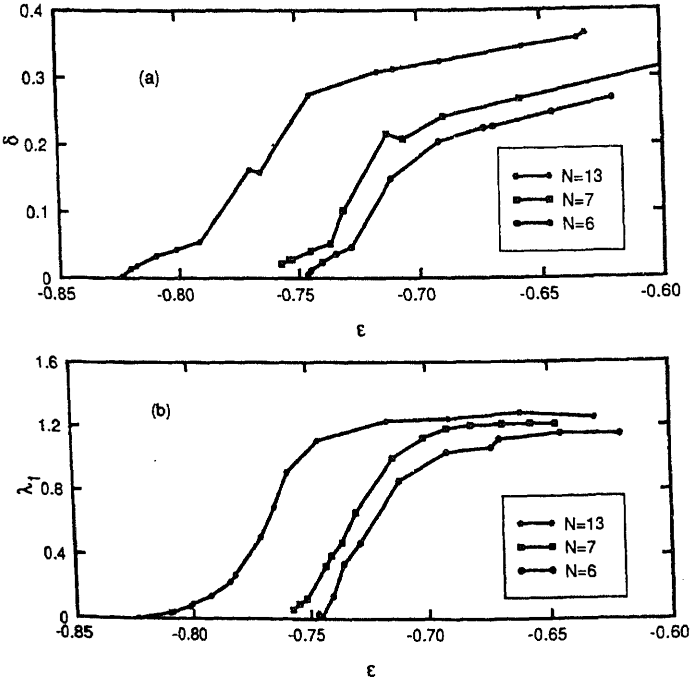


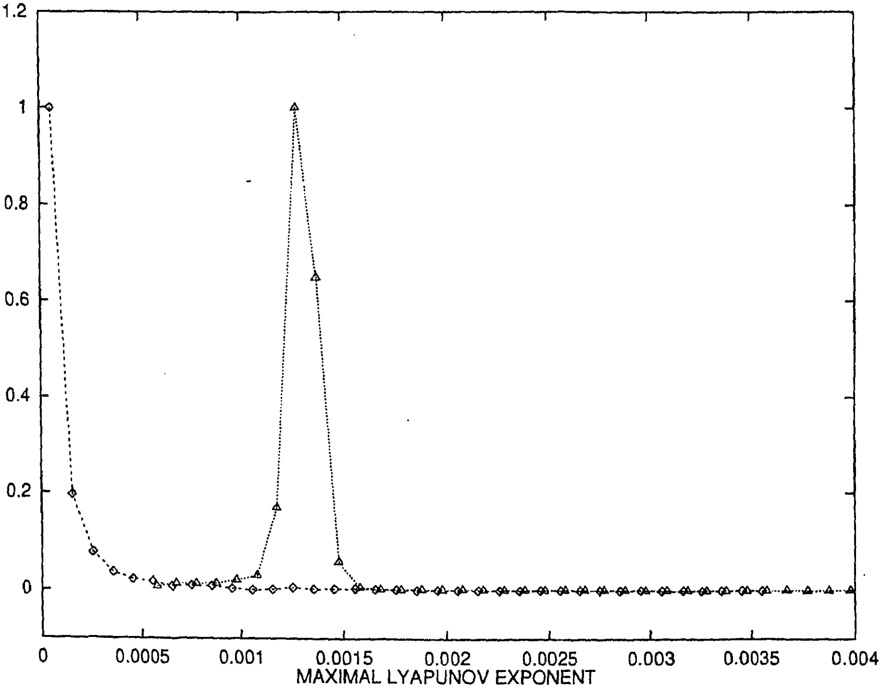
Figure 2. (**a**) Variation of Lindemann index with energy per particle (in reduced units) for Gupta$_N$ clusters for $N = 6, 7, 13$. (**b**) Variation of maximal Lyapunov exponent with energy per particle (in reduced units) for Gupta$_N$ clusters for $N = 6, 7, 13$.

observations and conclusions hold if one considers the behaviour of the MLE as a function of the average kinetic energy or temperature.

At the melting transition, both the Lindemann index and the Lyapunov exponent give identical information about the change in phase of the cluster. At other transitions, when a transition' is not as well defined, it is often difficult to find a reliable indicator of :hange. Take, for example, the putative glass-like transition that has been seen in :lusters, an example of which is provided by Ar_{19} [22]. The global behavior of the .as been examined [45] as a function of average kinetic energy or temperature for tarting from different configurations corresponding to the global minimum, local ı and saddles in the potential energy hypersurface. The radial function which is only used in bulk studies proves unreliable in detecting phase change, while the :an be shown to characterise the instability in a precise manner, both in bulk ıls as well as in finite clusters [45].

generality (and perhaps universality) of these observations needs to be investigated il. Mehra and Ramaswamy [46] have studied melting in transition-metal clusters he Gupta potential. In these systems there is no regime of solid-liquid coexistence, re can be a fluctuating state where the cluster continuously moves between two res with zero diffusion coefficient. Simulations of 6, 7, and 13 atom clusters have parallel behavior: no discontinuities were observed in the MLE, but a marked inuity in the slope near T_{m} was noted (figure 2).

her recent work on the dynamics at phase transitions, Bonasera *et al* [47] have d CMD simulations near the critical point for the liquid-gas transition. They study m of 100 nucleons interacting through a pairwise Yukawa potential, as well as a helium droplet containing about 800 He^4 atoms interacting via the Slater-ɔod potential and claim that the MLE has a peak at T_{c} and can be fitted near the



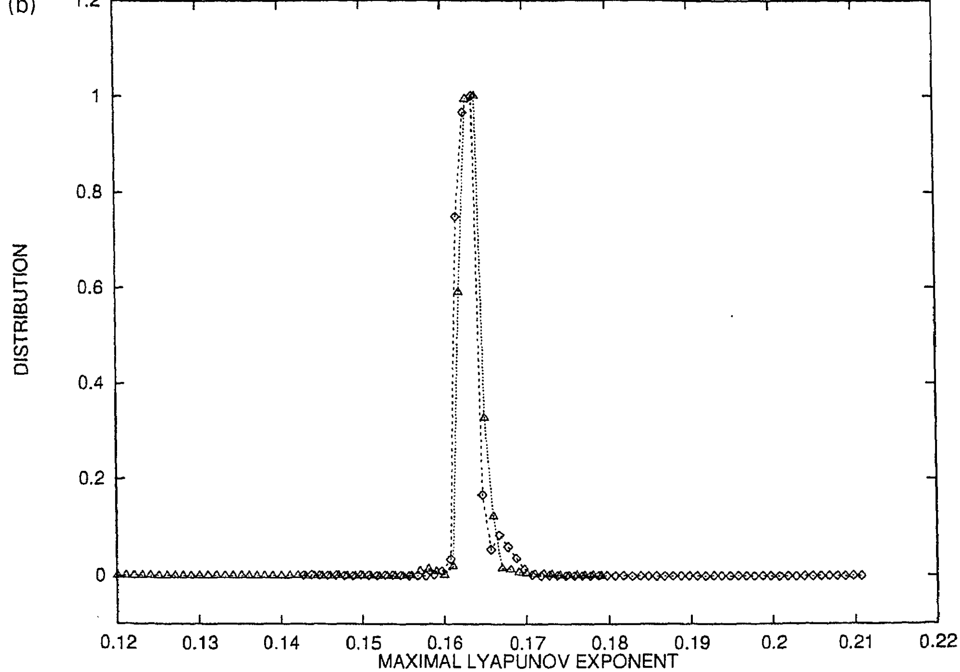


Figure 3. (**a**) Probability distribution of the maximal Lyapunov exponent for Ar_3. The two different symbols correspond to two different initial conditions. The distributions differ markedly and is a signature of non-ergodicity at low temperatures. See the text for details. (**b**) Probability distribution of MLE for Ar_3 for two different initial conditions at $E = -0.156 \times 10^{-13}$ erg/atom, $T = 32.3$ K. The distributions collapse on each other indicating that the system has attained ergodicity.

critical temperature with a power law

$$\lambda \sim |T - T_c|^{\omega}$$

with exponent $\omega = 0.15$ (in both the systems). It is not clear if this observation is on solid ground since, as has been pointed out by Nayak *et al* [47], the numerical fit to data is rather poor and the error bars in the data itself is unacceptably large. At continuous phase transitions though, the MLE shows characteristic and distinctive behaviour. Apart from the study of Butera and Caravati, there is recent work by Yamaguchi [10] who has studied a globally coupled Hamiltonian lattice which undergoes a continuous phase transition as energy per particle is increased. There is apparently a peak in the MLE at T_c although the model is integrable both in zero-energy and high-energy limits. It appears that at continuous phase transitions the Lyapunov exponent has a maximum value, while for first-order transitions MLE typically displays a sigmoidal curve. Welding these disparate observations into a coherent whole and giving a satisfactory explanation of the phenomena is currently a major theoretical challenge.

The general features of the behaviour of the MLE at phase transitions can be verified by studying simpler systems where the same behaviour occurs. Simple dissipative dynamical systems show so-called 'crises' [48], where the phase space volume changes discontinuously with the change in a system parameter. Crises (which also occur in conservative systems [49]) can be of several types. For attractor widening crises [50] where the volume increases by an order of magnitude or more, the MLE changes abruptly at the crisis point, increasing as a power law. For attractor merging crises when the phase space volume doubles, the variation of the MLE with parameter is "knee"-like, i.e. there is a change in the slope of the MLE. Interestingly, both these behaviours have been seen in the case of phase change in finite system (c.f. the cases of Ar_{13} and Ar_7), as well as in phase transitions in other systems.

Calculations of entire spectrum of Lyapunov exponents have also been made [36, 46, 51]. For the study of complex systems this is indeed important, but calculating them through the standard methods such as the tangent space technique can be computationally intensive. New methods proposed [52, 53] can circumvent this bottleneck. Lyapunov spectra characterize the short-time behaviour of correlation functions, and are related to the entire set of relaxation times in the system [54] although this has been established only for simple low-dimensional systems. For high-dimensional systems, in the limit $N \rightarrow \infty$, the existence of a smooth distribution λ_n, $n = i, \ldots, N$ has been numerically verified [55]. In the highly chaotic regime the function λ_n is linear [56], and this has been verified for Ar clusters (see figure 3(a)). In other systems, such as metallic clusters (using the Gupta potential) the results differ: the Lyapunov spectra for 6, 7 and 13 particle metal clusters have been calculated, and λ_n is found to be a nonlinear function of n [50] (figure 3(b)). The behaviour of Lyapunov spectra at phase transitions needs to be investigated.

4. Local K-entropy and local Lyapunov exponents

Global dynamical indicators are useful, but they are also somewhat limited in that they do not provide details about local dynamics—whether motion within a particular region is more or less chaotic. They also do not give any indication of the time-dependence of phenomena such as the spread of ergodicity. This information is provided by local indicators of the dynamics, namely the short time distributions of the K-entropy and the Lyapunov exponent. Berry and coworkers, in a series of papers [11, 36, 57], related the local K-entropy information to the topology of the cluster potential energy surface (PES). The local K-entropy is finite-time analogue of the global entropy, and is obtained by dividing a long trajectory into several parts of duration $N\Delta t$ and calculating the entropy for each time segment separately. From this the probability distribution of entropies is obtained, which can then be compared with the distribution of potential energy averaged over same duration $N\Delta t$.

A direct way to characterize the structure of a PES is to catalogue its various minima and saddles through steepest gradient and quenching techniques. This can be done for small clusters but as the number of minima and saddles increases exponentially with size, it is a formidable task to list all of these. The geometric features of a saddle can be determined by calculating the average curvature from the eigenvalues of the instantaneous Hessian matrix, namely the instantaneous normal mode spectra [24]. The analysis of clusters in saddle regions is important in understanding phenomena such as isomerization and solid-liquid transition.

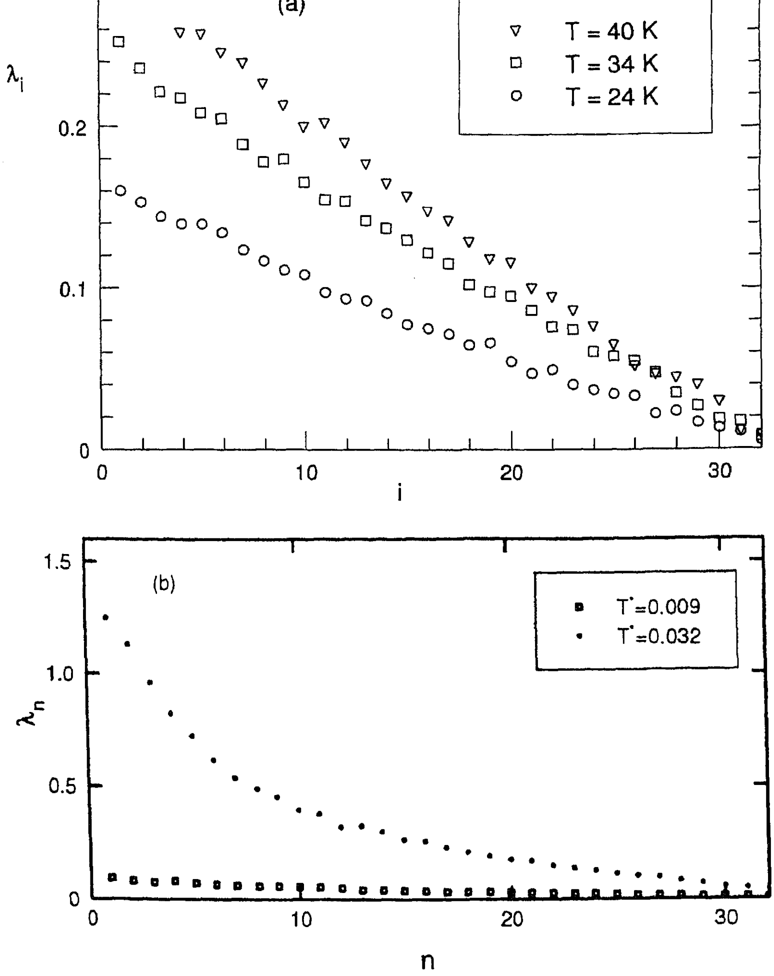


Figure 4. (**a**) Lyapunov spectra for Ar_{13} at three temperatures (1) 24 K (solid phase), (2) 34 K (coexistance phase) and (3) 40 K (liquid phase). (**b**) Lyapunov spectra for $Gupta_{13}$ at reduced temperatures (1) $T^* = 0.009$ (solid) and (2) $T^* = 0.032$ (liquid). Only the 32 positive exponents are shown.

In Ar_3 and Ar_4 clusters, which have rather flat saddles, the local K-entropy was found [11, 36, 57] to be significantly smaller in the saddle region as compared to other regions of the PES. This was attributed to an increase in the degree of quasiperiodicity near the saddle arising from a partial decoupling of the cluster's vibrational modes (as measured by the total vibrational coupling). In Ar_5 on the other hand, the local K-entropy in saddle regions was comparable to that in the well while for Ar_6 and Ar_7 saddle entropy was intermediate between entropies of the two wells which are connected by that saddle. All these clusters have relatively steep saddles, which suggests that the 'regularizing' effect of a saddle depends on the shape of PES near the saddle, especially the flatness of the saddle along the isomerization or reaction path. This behaviour is size-dependent: in larger clusters the influence of a particular saddle is reduced as the system will not spend much time on a sharp saddle.

In systems where the topography of the potential energy surface is very complicated, ergodicity cannot be guaranteed at all energies or temperatures. This can be seen by examining the distribution of finite-time or local Lyapunov exponents which are defined in the same manner as the local K-entropy, by dividing a long trajectory into segments of length $N\delta t$. Amitrano and Berry [57, 58] calculated the probability distribution of the largest local Lyapunov exponent for Ar_3 at different values of energy for various time intervals, and related these distributions to the evolution of ergodicity. At low energies ($T = 18\,K$) in the argon trimer, the shape of the distribution depends on the initial conditions for any time interval: the system does not exhibit ergodicity even on long time scales. At intermediate energies ($T = 28\,K$) the system is found to be ergodic: different initial conditions gave similar unimodal distributions but with different second moments. At energies high enough to allow saddle crossing ($T = 31\,K$) the distributions are bimodal up to a particular time interval, and unimodal for longer times, showing a time scale separation between the intra–well and inter–well motions. At still higher energy ($T = 37\,K$) one has true ergodic behavior with unimodal distributions even for shortest trajectory segments. Figure 3(a) is the probability distribution of MLE for Ar_3 at $E = -0.156 \times 10^{-13}$ erg/atom, $T = 7.41\,K$ for two different initial conditions, the trajectories being of 10^8 MD time steps. The distributions shows no sign of convergence, indicating the non-ergodic nature of the system at that temperature. In contrast, the probability distributions of MLE at a higher energy $E = -0.089 \times 10^{-13}$ erg/atom, $T = 32.3\,K$ (see figure 3(b)) for two different initial conditions collapse on each other indicating the system has attained ergodic behavior. These observations at the microscopic level are significant particularly in the context of transition from non-ergodic to ergodic behavior in small finite systems. Similarly, at $T = 14\,K$ the autocorrelation function of the local Lyapunov function is oscillatory, while it rapidly decays to zero for $T \geq 18\,K$. Qualitatively similar results hold for Ar_7. A general theory of these phenomena has not yet been given, and it is not a simple matter to extend these calculations to larger cluster sizes.

5. Discussion and summary

Finite cluster systems may provide one regime where the methods of dynamical systems and statistical mechanics can simultaneously be applied. Given the highly nonlinear character of the typical cluster dynamical system, it is not surprising that chaotic motion is both so widespread and that it plays such an important role. For small clusters, although the number of freedoms is not too large, it is possible to focus on particular features of interest—such as the K-entropy or the Lyapunov exponents—and explore the dynamics of how statistical behaviour such as ergodicity develops in such systems.

There are, however, dynamical regimes of interest when the motion of clusters becomes simple. For example, at very low energies above the global minimum, the cluster, like many hamiltonian systems, can be expected to be nearly integrable. In a recent study of Ar_{13} clusters, Salian *et al* [59] examined the dynamics of collective excitations by exciting the monopole mode, namely that mode of oscillation where the particles execute radial motion only, as a function of temperature. Below a threshold, which they determined to be $T_s = 7.2$ K, this mode does not decay for long times. Above T_s, the time dependence of the amplitude variation could be accurately modeled by a damped harmonic oscillator, while below T_s, the monopole mode is periodic or quasiperiodic. This suggests [59] that as long as the cluster is confined to the global minimum in the PES, the monopole mode does not decay. When the temperature is high enough for the system to come out of this well, the nearby local minima, which correspond to rearrangement of one or a few particles act as decay channels, akin to single–particle excitations in nuclei [60]. The MLE in this temperature range has also been studied [61]: below T_s, the MLE is essentially zero, and although in the so-called 'solid' state, the MLE becomes positive only above T_s. Thus the survival of the symmetric monopole mode is intimately connected to the onset of globally chaotic motions in the dynamics.

A major focus of this article has been on the appropriateness of using the Lyapunov exponents as indicators of phase change since this focuses on a property of the phase space behaviour of the system. The dynamics of the solid-liquid melting transition, the liquid-gas transition, and the glass transition in finite systems have been studied using this quantity. For phase transitions where order parameters based on other criteria can be ambiguous (as for example the glass transition), the MLE may be able to provide the clearest signature of the transition.

References

[1] J Lebowitz and O Penrose, *Phys. Today*, February 1973, p. 23

[2] See, *e.g. The physics and chemistry of small clusters*, edited by P Jena, B K Rao and S N Khanna, (New York, Plenum, 1987)
Physics and Chemistry of finite systems: From clusters to crystals, edited by P Jena, S N Khanna and B K Rao, (Kluwer Dordrecht, 1992)
Clusters of atoms and molecules, edited by H. Haberland (Springer-Verlag, 1994)

[3] R S Berry, in *Clusters of atoms and molecules*, edited by H Haberland (Springer-Verlag, 1994)

[4] S G Kim and D Tománek, *Phys. Rev. Lett.* **72**, 2418 (1994)

[5] See *e.g.* R F Service, *Science* **271**, 920 (1996)

[6] D J Wales, *Science* **271**, 925 (1996)

[7] H Haberland, in *Clusters of atoms and molecules*, edited by H Haberland (Springer-Verlag, 1994)

[8] S Sugano, *Microcluster Physics*, (Springer, Berlin, 1991)

[9] M Pettini, *Phys. Rev.* **E 47**, 828 (1991)

[10] Y Y Yamaguchi, *Slow relaxation at critical point of second order phase transition in a highly chaotic hamiltonian system*, LANL archives chao-dyn 9602002

[11] R J Hinde and R S Berry, *J. Chem. Phys.* **99**, 2942 (1993)

[12] J K Lee, J A Barker and F F Abraham, *J. Chem. Phys.* **58**, 3166 (1973)

[13] C L Briant and J J Burton, *Nature Physical Science* **243**, 100 (1973); *J. Chem. Phys.* **63**, 2045 (1975)

[14] *Elemental and molecular clusters*, edited by G Benedek, T P Martin and G Pacchioni, (Springer, Berlin, 1988)

[15] A Heidenreich, I Oref and J Jortner, *J. Chem. Phys.* **96**, 7517 (1992)

[16] R S Berry, T L Beck, H L Davis and J Jellinek, *Adv. Chem. Phys.* **70**, 75 (1988)

[17] R S Berry, *Sci. Am.* **61**, 56 (1990)

[18] D Wales, *Mol. Phys.* **78**, 151 (1993)

[19] H Matsuoka, T Hirokawa, M Matsui and M Doyama, *Phys. Rev. Lett.* **69**, 297 (1992)
[20] H P Cheng and R S Berry, in *Symposium on clusters and cluster-assembled materials*, Materials Research Society, pp 241–252, (1991); H P Cheng and R S Berry, *Phys. Rev.* **A 45**, 7969 (1992)
[21] J-B Maillet, A Boutin and A H Fuchs, *Phys. Rev. Lett.* **76**, 4336 (1996)
[22] K D Ball, R S Berry, R E Kunz, F Li, A Proykova, and D J Wales, *Science* **271**, 963 (1996)
[23] M P Allen and D J Tildesley, *Computer simulation of liquids* (Oxford University Press, Oxford, 1987)
[24] See *e.g.* R. M. Stratt, Acc. Chem. Res. **28**, 201 (1995) A technique to compute Lyapunov exponents from equilibrium simulations is discussed by C. Chakravarty and R. Ramaswamy in *Instantaneous normal mode spectra of quantum clusters, J. Chem. Phys.* (in press)
[25] R P Gupta, *Phys. Rev.* **B23**, 6265 (1981)
[26] D Tomanek, S Mukherjee and K H Bennemann, *Phys. Rev.* **B28**, 665
[27] S Sawada and S Sugano, *Z. Phys.* **D14**, 247 (1989)
[28] N Ju and A Bulgac, *Phys. Rev.* **B48**, 2721 (1993)
[29] T Kanishi and K Kaneko, *J. Phys.* **A25**, 6283 (1992)
[30] H D Meyer, *J. Chem. Phys.* **84**, 3147 (1986)
[31] Ya B Pesin, *Russ. Math. Surveys,* **32**, 55 (1977)
[32] G Benettin, L Galgani and J M Strelcyn, *Phys. Rev.* **A14**, 2338 (1976)
[33] J P Eckmann, and D Ruelle, *Rev. Mod. Phys.*, **57**, 617 (1985)
[34] P Butera and G Caravati, *Phys. Rev.* **A36**, 962 (1987)
[35] H A Posch and W G Hoover, *Phys. Rev.* **A38**, 473 (1988); *Phys. Rev.* **A39**, 2175 (1989)
[36] R J Hinde, R S Berry and D J Wales, *J. Chem. Phys.* **96**, 1376 (1992)
[37] S K Nayak, R Ramaswamy and C Chakravarty, *Phys. Rev.* **E51**, 3376 (1995)
[38] F A Lindemann, *Phys. Zeits.* **14**, 600 (1910)
[39] S K Ma, *Statistical physics*, (World Scientific, Singapore, 1985)
[40] D J Wales and R S Berry, *J. Chem. Phys.* **92**, 4283 (1990)
[41] T L Beck, J Jellinek and R S Berry, *J. Chem. Phys.* **87**, 545 (1987)
[42] H L Davis, J Jellinek and R S Berry, *J. Chem. Phys.* **86**, 6456 (1987)
[43] H A Posch and W G Hoover, *Phys. Rev.* **A38**, 473 (1988); *Phys. Rev.* **A39**, 2175 (1989)
[44] P Labastie and R L Whetten, *Phys. Rev. Lett.* **65**, 1567 (1990)
[45] S K Nayak, P Jena, K D Ball and R S Berry, to be published
[46] V Mehra and R Ramaswamy, to be published
[47] A Bonasera, V Latora and A Rapisarda, *Phys. Rev. Lett.* **75**, 3434 (1995)
S K Nayak, P Jena, A Bhattacharya and S D Mahanti, *Phys. Rev. Lett.* **00**, (submitted, 1996)
[48] V Mehra and R Ramaswamy, *Phys. Rev.* **E53**, 3420 (1996)
[49] M Toda, *Phys. Rev. Lett.* **74**, 2670 (1995)
[50] C Grebogi, E Ott, J Romeiras and J Yorke, *Phys. Rev.* **A36**, 5365 (1987)
[51] M V Kuzmin, I V Nemov, A A Stuchebrukhov, V N Bagratashvili and V S Letokhov, *Chem. Phys. Lett.* **124**, 522 (1986)
[52] S Sastry, *Phys. Rev. Lett.* **76**, 3738 (1996)
[53] S Habib and O Pyne, *Phys. Rev. Lett.* **74**, 70 (1995)
[54] G M Zaslavskii, *Chaos in dynamical systems* (Harwood Academic, 1985)
[55] R Livi, M Politi, S Ruffo and A Vulpiani, *J. Stat. Phys.* **46**, 147 (1987)
[56] G Paladin and A Vulpiani, *Phys. Lett.* **A118**, 14 (1986)
[57] C Amitrano and R S Berry, *Phys. Rev. Lett.* **68**, 729 (1992)
[58] C Amitrano and R S Berry, *Z. Phys.* **D26**, 388 (1993)
[59] U A Salian, S N Behera and V S Ramamurthy, *J. Chem. Phys.* **105**, 3679 (1996)
[60] S J Wang and W Cassing, *Ann. Phys. (N.Y.)* **159**, 328 (1985)
W Cassing and A Pfitzner, *Z. Phys.* **A337**, 175 (1990)
[61] V Mehra and R Ramaswamy, unpublished

Classical gauge theories as dynamical systems—Regularity and chaos

S LAKSHMIBALA, BINDU A BAMBAH[1], M S SRIRAM[2] and C MUKKU[3]
Department of Physics, Indian Institute of Technology, Madras 600 036, India
[1]School of Physics, University of Hyderabad, Hyderabad 500 046, India
[2]Department of Theoretical Physics, University of Madras, Madras 600 025, India
[3]School of Mathematics and Computer/Information Sciences, University of Hyderabad, Hyderabad 500 046, India

Abstract. In this review we present the salient features of dynamical chaos in classical gauge theories with spatially homogeneous fields. The chaotic behaviour displayed by both abelian and non-abelian gauge theories and the effect of the Higgs term in both cases are discussed. The role of the Chern–Simons term in these theories is examined in detail. Whereas, in the abelian case, the pure Chern–Simons–Higgs system is integrable, addition of the Maxwell term renders the system chaotic. In contrast, the non-abelian Chern–Simons–Higgs system is chaotic both in the presence and the absence of the Yang–Mills term. We support our conclusions with numerical studies on plots of phase trajectories and Lyapunov exponents. Analytical tests of integrability such as the Painlevé criterion are carried out for these theories. The role of the various terms in the Hamiltonians for the abelian Higgs, Yang–Mills–Higgs and Yang–Mills–Chern–Simons–Higgs systems with spatially homogeneous fields, in determining the nature of order-disorder transitions is highlighted, and the effects are shown to be counter-intuitive in the last-named system.

Keywords. Gauge theories; integrability; chaos; phase space; Lyapunov exponents.

PACS Nos 11.15; 05.45

1. Introduction and overview

The dynamical behaviour of gauge theories, whose importance in particle physics is well known, has been studied in detail in recent times. This has been a primer to the understanding of the chaotic behaviour of field theories. Early investigations along these lines in the context of non-abelian gauge theories were carried out with an aim to understanding the structure of the field-theoretic vacuum and the asymptotic states of the theory [1] with particular reference to strong interaction physics. This was motivated by the belief that there was a connection between colour confinement and chaos in quantum chromodynamics, the popular theory of strong interactions. Further interest in such investigations arose because some gauge theories support solitons, and in a sense, these topological structures and chaos are examples of opposite extremes of nonlinear behaviour, and their coexistence in field theories needs to be examined and understood.

Since the systems considered are conservative and are described by Hamiltonians, non-integrability of the evolution equations implies chaos. From this point of view, extensive studies have been conducted on the classical Yang–Mills (YM) system which is the non-abelian gauge theory that (when suitably adapted) describes certain fundamental interactions in particle physics. Even in the late seventies, attention was drawn to the fact that only if self-duality conditions are imposed, are the classical YM equations integrable [2]. This was confirmed by the Painlevé test for integrability in the special case of the self-dual $SU(2)$ YM theory [3]. In the general case, where self-duality conditions are not imposed, most results reported in the literature pertain to a special reduction of the classical YM equation. This reduction corresponds to obtaining a dynamical system (DS) from the field theory by considering the fields to be spatially homogeneous.

Although the DS is an extremely simplified version of any field theory, there exists a well-defined limit where this simplification is justified. The field configurations relevant for the ground state of quantum chromodynamics, for instance, would vary only slowly in space and would correspond to a long wavelength region. The DS involving spatially homogeneous fields is a good approximation in this case. In particle production it is conjectured that chaotic field configurations may be excited during violent collisions of particles, in localized regions of space, where the fields are approximately homogeneous. These could contribute significantly to some physical processes, in which their signatures would be unique [4].

The DS that is obtained from the field theory by assuming the fields to be spatially homogeneous is clearly governed by coupled ordinary differential equations describing the time evolution of the fields, in contrast to the full field theory described by a complicated set of coupled partial differential equations. YMDS which is the dynamical system obtained in this manner from the YM field theory, was shown numerically to be chaotic, in the $SU(2)$ case, for a suitable choice of ansatz for the spatially homogeneous gauge fields [5]. The ansatz reduces the theory to a dynamical system with two degrees of freedom, since the time component A_0^a of the gauge potential (a is an $SU(2)$ index) is set equal to zero and the space components are defined in terms of two variables alone. Whereas the Liouville-Arnol'd criterion for integrability requires the existence of two constants of the motion, only the Hamiltonian corresponding to YMDS is conserved and no additional first integral exists. Numerical studies of the phase trajectories add support to the inference that YMDS is chaotic. In [5] it has been conjectured that since there exists at least one reduction of the YM field theory which is chaotic, the full field theory should also exhibit chaos. It is worth noting that yet another reduction obtained by a different choice of ansatz for the spatially homogeneous fields leads to nonlinear plane wave solutions of the YM equations in Minkowski space, which also exhibit chaotic behaviour [6]. Analytical tests of integrability like the ARS method [7] would be conducted in the same spirit, by examining in principle all reductions of the coupled partial differential equations describing the field theory of interest. The complete field theory is said to be integrable if all reductions pass the analytical test for integrability.

The YMDS equation has been investigated from the point of view of ergodic theory [8, 9] wherein it is inferred that it displays strong statistical properties, and is homeomorphic to the Krylov–Sinai billiards system with hyperbolic walls.

The chaotic nature of the YM field equations has been explicitly verified in [10, 11] where it is shown that the phase space near the Wu–Yang static solution is a stochastic region. The verification of the existence of spatio-temporal chaos in this case, relies essentially on the Fermi–Pasta–Ulam approach [12] where the continuous string is replaced by a discrete lattice, for numerical purposes.

These conclusions based on YMDS (and further strengthened by explicit field-theoretic computations) change significantly when Higgs terms are added to the gauge field. The Yang–Mills–Higgs (YMH) system has been extensively investigated in this respect. YMH is important in particle physics as several physical models based on it, incorporating spontaneous breakdown of the gauge symmetry, have been suggested and successfully applied to the unification of the fundamental forces of nature.

From the point of view of DS, the $SO(3)$ Georgi–Glashow model which admits monopole solutions has been shown (under suitable assumptions), to exhibit chaos when reduced to its dynamical counterpart where the fields are considered to be spatially homogeneous [13, 14]. By applying the Painlevé prescription it has been established in [14] that the model, with a suitable ansatz for the gauge fields and the Higgs field ϕ, is algebraically non-integrable for $\lambda > 2g^2$. Here λ is the coefficient of ϕ^4 in the Lagrangian and g is the gauge coupling constant. The Toda–Brumer criterion [15] when applied to this model yields an expression for the critical energy for the onset of chaos, in terms of λ and m (the coefficient of ϕ^2 in the Lagrangian).The full field-theoretic version of the $SO(3)$ Georgi–Glashow model has also been investigated by considering the phase space around the t'Hooft–Polyakov monopole solution [16]. There exists a phase transition in the system in the following sense: For small perturbations around the t'Hooft–Polyakov monopole solution, the motion is regular and the system is close to an integrable one. As the strength of the perturbation increases above a critical value, the motion becomes chaotic. Since the pure YM system is always chaotic, this result shows that the Higgs fields are necessary to have an ordered phase in the theory.

In order to address the question of whether the order-chaos transition is an inherent phenomenon only observed in the case where a monopole solution exists in the system, a different model has been considered in the literature [17]. This is described by an $SU(2)$ YMH theory which admits spherically symmetric sphaleron solutions and where the fields are taken to be spatially homogeneous. Through extensive numerical work it is established in this case too that an order-chaos transition occurs similar to that observed in the monopole system.

The nature of chaos that arises in the abelian counterparts of the foregoing has also been investigated [13, 14, 18]. The abelian Higgs (AH) model as a DS is shown to be algebraically non-integrable in most of the type-I region of superconductivity [14]. The conclusion is further supported by numerical work including investigation of phase plots and Lyapunov exponents in both the Georgi–Glashow and AH cases [13].

It therefore appears from the details reviewed above that a chaotic system like YMDS exhibits regularity for some parameter values if the Higgs term is introduced and that a transition from order to chaos takes place in the YMH system. However, in contrast, the message from the abelian counterpart seems to be that a regular system (Maxwell) undergoes an order-to-chaos transition on introduction of the Higgs field. At this stage, therefore, the role of the Higgs term is still not clear. More detailed studies on this aspect have been undertaken in [19, 20, 21].

A related question pertains to whether other terms apart from the Higgs coupling can produce order in YMDS for any suitable range of the parameters of the theory. A possible candidate in this regard would be the Chern–Simons (CS) term. In $(2+1)$ dimensions the CS term plays a very interesting role in both abelian electrodynamics and non-abelian gauge theories [see, for example, [22]]. Owing to its presence, the gauge field acquires mass and yet the theory is gauge invariant. The CS term can be generated through spontaneous symmetry breaking [23]. Since it seems to implement the effect of the Higgs term in this sense, it is important to know its effect on YM from the dynamical systems point of view as well.

Interesting effects have been observed [4] in the YMCS system when the fields are made spatially homogeneous. Order-chaos transitions occur within narrow energy ranges and have been tracked in this model, for instance, in the energy range between 0.37 and 0.40 in suitable units. Below the lower limit of energy the system displays order and above the upper limit, chaos. The authors have observed a hyperfine structure of order-chaos-order-$\cdots$ transitions present at all arbitrary small scales within the energy range 0.37 to 0.40.

Thus while both the CS and Higgs terms seem to bring about order-chaos transitions in YMDS (which is otherwise totally chaotic), the transition region seems to display different stochastic properties depending on whether the CS term or the Higgs term is added to the gauge theory. It is therefore of interest to explore in detail the combined effect of both the CS and Higgs couplings in the abelian and non-abelian situations. New features emerge in this case, some of them rather counter-intuitive. The details of these calculations are reported in [19, 20, 21], and will form the contents of the following sections. Here we give a summary of the results.

We find that the abelian Chern–Simons–Higgs system without the Maxwell term is integrable, whereas the inclusion of the Maxwell term makes the system chaotic [19]. We then consider the $SU(2)$ Yang–Mills–Chern–Simons–Higgs system with homogeneous fields, both in the absence of the YM term (NACSH) and in its presence (YMCSH) [20]. In contrast to the abelian case, where, in the absence of the Maxwell term, the system is integrable, here both NACSH and YMCSH are non-integrable. All this is verified by a Painlevé analysis as well as by a study of phase plots and Lyapunov exponents. In the case of YMCSH, the dimensionality of the phase space being large (eighteen), we have restricted ourselves to those sections of phase space where we expect interesting features and our plots reveal that the dynamical evolution ranges from regular to chaotic behaviour. We see that in the limit of large Chern–Simons mass, large Higgs coupling constant, or small gauge coupling constant, NACSH and YMCSH become regular.

In the above analysis, we retain the full vector nature of the gauge fields and do not restrict our study to just a single ansatz, in contrast to some of the earlier works reported in the literature, on classical gauge theories as dynamical systems. We have examined in detail the Lyapunov spectra for NACSH and YMCSH as functions of the various couplings and parameters in the system [21]. We find that there exists an interplay between the oscillatory effects of the CS term, the regularising effect of the Higgs term and the completely chaotic nature of the YM term. However, a striking feature that emerges in our investigations is that it is not true that an increase in the Higgs coupling constant regularizes the gauge term at all energies. This only happens, in fact, for small values of the energy; in this sense, the result is counter-intuitive.

2. The abelian Chern–Simons–Higgs theory

2.1 *The dynamical system*

The Lagrangian density of the $(2+1)$ dimensional Chern–Simons–Higgs system is given by

$$L_0 = \frac{m}{2}\epsilon^{\mu\nu\rho}A_\mu F_{\nu\rho} + |(\partial_\mu - ieA_\mu)\phi|^2 - V(\phi), \tag{1}$$

where $V(\phi)$ is the Higgs potential and m is the Chern–Simons mass. The equations of motion are

$$\frac{1}{2}\epsilon_{\mu\nu\rho}F^{\nu\rho} = \frac{-ie}{2m}(\phi^*\partial_\mu\phi - \phi\partial_\mu\phi^*) - \frac{e^2}{m}A_\mu|\phi|^2, \tag{2}$$

$$(\partial_\mu - ieA_\mu)(\partial^\mu - ieA^\mu)\phi = -\frac{\partial V}{\partial\phi^*}. \tag{3}$$

The corresponding dynamical system (CSH) is obtained (as described in the preceding section) by taking the fields to be spatially homogeneous. We also choose the gauge $A_0 = 0$, for convenience. Further, it can be easily seen that the phase of ϕ is a constant. Thus, ϕ can be taken to be real without loss of generality.

The equations of motion for A_1, A_2 and ϕ then become

$$\dot{A}_2 = -\frac{e^2}{m}A_1\phi^2, \tag{4}$$

$$\dot{A}_1 = \frac{e^2}{m}A_2\phi^2, \tag{5}$$

$$\ddot{\phi} + e^2({A_1}^2 + {A_2}^2)\phi = -\frac{1}{2}\frac{\partial V}{\partial\phi}. \tag{6}$$

The overhead dot refers to differentiation with respect to time. From these equations it follows that $({A_1}^2 + {A_2}^2)$ is a constant of the motion. Writing $A_1 = C\cos\chi$ and $A_2 = C\sin\chi$, we get

$$\dot{\chi} = -\frac{e^2}{m}\phi^2, \tag{7}$$

$$\ddot{\phi} + e^2C^2\phi = -\frac{1}{2}\frac{\partial V}{\partial\phi}. \tag{8}$$

Hence the Chern–Simons–Higgs field theory reduces to a dynamical system, with two degrees of freedom ϕ and χ whose equations of motion are as given above.

The Lagrangian which yields these equations is

$$L_c = p_\chi\left[\dot{\chi} + \frac{e^2}{m}\phi^2\right] + (\dot{\phi})^2 - e^2C'^2\phi^2 - V(\phi), \tag{9}$$

where p_χ is a Lagrange multiplier, and also the canonical momentum corresponding to χ. Also,

$$C^2 = C'^2 - \frac{p_\chi}{m}. \tag{10}$$

The corresponding Hamiltonian is

$$H = \frac{p_\phi^2}{4} - \frac{e^2}{m} p_\chi \phi^2 + e^2 C'^2 \phi^2 + V(\phi) \tag{11}$$

so that the equations of motion are

$$\dot{\phi} = \frac{\partial H}{\partial p_\phi} = \frac{p_\phi}{2}, \tag{12}$$

$$\dot{p_\phi} = -\frac{\partial H}{\partial \phi} = \frac{2e^2}{m} p_\chi \phi - 2e^2 C'^2 \phi = -2e^2 C^2 \phi - \frac{\partial V}{\partial \phi}, \tag{13}$$

$$\dot{\chi} = \frac{\partial H}{\partial p_\chi} = -\frac{e^2}{m}\phi^2, \tag{14}$$

$$\dot{p_\chi} = -\frac{\partial H}{\partial \chi} = 0. \tag{15}$$

Thus we see that p_χ is a constant of the motion. Hence CSH has two degrees of freedom, and two integrals of the motion, H and p_χ. Therefore the Liouville–Arnol'd criterion for integrability is satisfied by this DS. The Painlevé test for integrability can be carried out trivially and confirms that CSH is indeed an integrable system.

2.2 *Maxwell–Chern–Simons–Higgs theory*

We now turn our attention to the situation when the Maxwell term is added to the Chern–Simons–Higgs field theory. The corresponding Lagrangian is given by

$$L = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{m}{2} \epsilon^{\mu\nu\rho} A_\mu F_{\nu\rho} + |(\partial_\mu - ieA_\mu)\phi|^2 - V(\phi). \tag{16}$$

Again, choosing the gauge $A_0 = 0$, and working with spatially homogeneous fields, we see that ϕ can be taken to be real, without loss of generality. The equations of motion describing this DS, which we will refer to as MCSH, are as follows:

$$\ddot{A}_1 + 2m\dot{A}_2 + 2e^2 A_1 \phi^2 = 0, \tag{17}$$

$$\ddot{A}_2 - 2m\dot{A}_1 + 2e^2 A_2 \phi^2 = 0, \tag{18}$$

$$\ddot{\phi} + e^2({A_1}^2 + {A_2}^2)\phi = -\frac{1}{2}\frac{\partial V}{\partial \phi}. \tag{19}$$

This system clearly has three degrees of freedom, A_1, A_2 and ϕ. The Hamiltonian appropriate to this system is given by

$$H = \tfrac{1}{2}[(p_1 - mA_2)^2 + (p_2 + mA_1)^2] + \frac{{p_3}^2}{4} + e^2({A_1}^2 + {A_2}^2)\phi^2 + V(\phi), \tag{20}$$

where p_1, p_2, and p_3 are the canonical momenta corresponding to A_1, A_2 and ϕ respectively.

We work with the quartic Higgs potential

$$V(\phi) = \frac{\lambda}{4}(\phi^2 - v^2)^2 \tag{21}$$

and rename A_1 as q_1, A_2 as q_2, and ϕ as q_3. Hamilton's equations then become

$$\dot{q}_1 = p_1 - mq_2, \tag{22}$$

$$\dot{q}_2 = p_2 + mq_1, \tag{23}$$

$$\dot{q}_3 = \frac{p_3}{2}, \tag{24}$$

$$\dot{p}_1 = -m(p_2 + mq_1) - 2e^2 q_1 {q_3}^2, \tag{25}$$

$$\dot{p}_2 = m(p_1 - mq_2) - 2e^2 q_2 {q_3}^2, \tag{26}$$

$$\dot{p}_3 = -2e^2(q_1^2 + q_2^2)q_3 - \lambda q_3(q_3^2 - v^2). \tag{27}$$

We note that corresponding to the MCSH Hamiltonian, $p_\zeta = A_2 p_1 - A_1 p_2$, is a constant of the motion. This feature can be used to express MCSH in terms of just two degrees of freedom. In close analogy with the CSH case, we now define $A_1 = A \cos \zeta$ and $A_2 = A \sin \zeta$. In terms of A, ϕ and ζ, Hamilton's equations read

$$\dot{A} = p_A, \tag{28}$$

$$\dot{p}_A = -2e^2\phi^2 A + \frac{(p_\zeta)^2}{A^3} - m^2 A, \tag{29}$$

$$\dot{\phi} = \frac{p_\phi}{2}, \tag{30}$$

$$\dot{p}_\phi = -2e^2 A^2 \phi - \lambda\phi(\phi^2 - v^2), \tag{31}$$

$$\dot{\zeta} = m - \frac{{p_\zeta}^2}{A^3}, \tag{32}$$

$$\dot{p}_\zeta = 0. \tag{33}$$

This system has three degrees of freedom, while there appear to exist only two constants of the motion. Hence one may suspect that the system could be chaotic. In order to verify this, the Painlevé test is sketched below. We first recast the above equations in the form of the following second-order differential equations by exploiting the fact that ζ is a cyclic coordinate. We have

$$\ddot{\phi} = -e^2\phi A^2 - \frac{\lambda}{2}\phi(\phi^2 - v^2) \tag{34}$$

and

$$\ddot{A} = -m^2 A + \frac{{p_\zeta}^2}{A^3} - 2e^2 A^2 \phi^2. \tag{35}$$

In terms of the rescaled variables $x = \phi/v$, $y = A/v$, $t' = evt$, $\mathcal{C} = {p_\zeta}^2/e^2 v^6$, $\mathcal{A} = \lambda/2e^2$, and $\mathcal{B} = m^2/e^2 v^2$, we get the equations

$$\ddot{x} = -xy^2 - \mathcal{A}x(x^2 - 1), \tag{36}$$

$$\ddot{y} = -\mathcal{B}y + \frac{\mathcal{C}}{y^3} - 2x^2 y, \tag{37}$$

where the differentiation is with respect to t'.

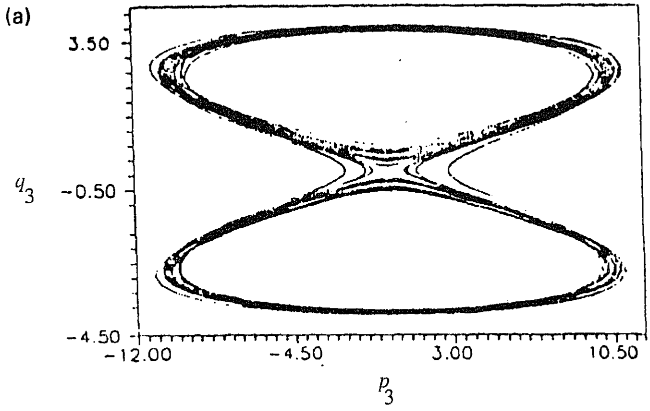


Figure 1(a). Plot of q_3 vs p_3 for the Maxwell Chern–Simons–Higgs system with parameters $m = 2, \lambda = 2, v^2 = 8$ and $e^2 = 1$ and initial conditions $q_1 = 1, q_2 = 1$, $q_3 = 0.5, p_1 = 1, p_2 = 1$ and $p_3 = 1$.

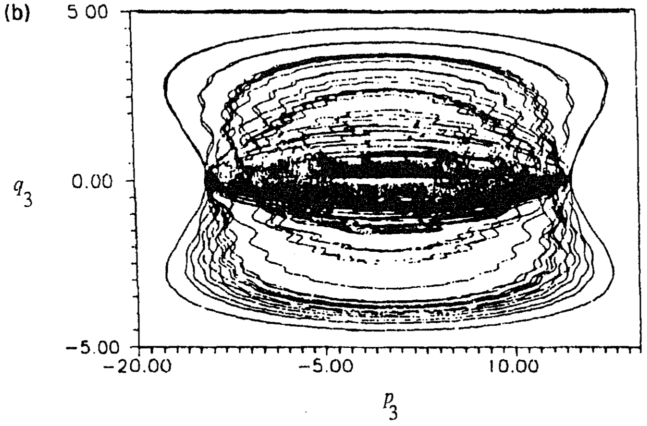


Figure 1(b). Plot of q_3 vs p_3 for the Maxwell Chern–Simons–Higgs system with parameters $m = 2, \lambda = 2, v^2 = 8$ and $e^2 = 100$ and initial conditions $q_1 = 1, q_2 = 1$, $q_3 = 0.5, p_1 = 1, p_2 = 1$ and $p_3 = 1$.

We first find the dominant singularity structure by setting $x = a(t - t_0)^\alpha$ and $y = b(t - t_0)^\beta$, where α and β must be negative. This reveals that there exist two possibilities for the leading behaviour: *Case* (1): $\alpha = -1, \beta = -1, a^2 = -1$ and $b^2 = \mathcal{A} - 2$. *Case* (2): $\alpha = -1, \beta = \frac{1}{2} \pm \frac{1}{2}\sqrt{(1 + (16/\mathcal{A}))}$, $a^2 = -2/\mathcal{A}$, and b is arbitrary. For both cases, the resonance analysis reveals that the Painlevé property fails for all physical regions of the Higgs potential, thus establishing that MCSH is not integrable.

2.3 *Numerical studies*

To establish the chaotic behaviour of the MCSH system numerically, we report below the results of a study [19] of the corresponding phase space trajectories. In figures 1(a) and

(b) we have displayed phase plots relating to MCSH. These are plots of q_3 vs p_3 for initial conditions $q_1 = 1$, $q_2 = 1$, $q_3 = 0.5$, $p_1 = 1$, $p_2 = 1$ and $p_3 = 1$. The parameter e^2 is chosen differently in figures 1(a) and (b) as seen in the corresponding captions. Further, while the CSH system is integrable, the MCSH plots are space-filling, signalling that MCSH is not. Also, it is clear from figures 1(a) and (b) that with increase in the value of e^2, MCSH becomes more space-filling. A study of the maximal Lyapunov exponent for MCSH as a function of the dimensionless parameter $\lambda/2e^2$ reveals that it is positive, thus establishing firmly that the system is chaotic.

3. The non-abelian Chern–Simons–Higgs theory

3.1 *The pure Chern–Simons–Higgs system*

The Lagrangian for the non-abelian $SU(2)$ Chern–Simons–Higgs field theory in $(2+1)$ dimensional Minkowski space is given by

$$L = \frac{m}{2}\epsilon^{\mu\nu\alpha}\left[F^a_{\mu\nu}A^a_\alpha - \frac{g}{3}\epsilon_{abc}A^a_\mu A^b_\nu A^c_\alpha\right] + D_\mu\phi^\dagger_a D^\mu\phi_a - V(|\phi|) \tag{38}$$

where

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\epsilon_{abc}A^b_\mu A^c_\nu , \tag{39}$$

and

$$D_\mu\phi_a = (\partial_\mu - igT^l A^l_\mu)\phi_a . \tag{40}$$

Here T_a's are the generators of the $SU(2)$ algebra, and $\mathrm{tr}[T_a T_b] = 2\delta_{ab}$. We consider a real triplet representation for the Higgs field, the gauge $\mathbf{A}_0 = 0$, and spatially homogeneous fields as before. Bold face letters denote vectors in $SU(2)$ space. Corresponding to this mechanical system (NACSH) the Gauss' law constraint is as given below:

$$m\mathbf{A}_1 \times \mathbf{A}_2 = -\mathbf{\Phi} \times \dot{\mathbf{\Phi}}. \tag{41}$$

The remaining equations of motion for the vector field are

$$\dot{\mathbf{A}}_1 = \frac{g^2}{m}(\mathbf{A}_2\Phi^2 - \mathbf{\Phi}\mathbf{A}_2 \cdot \mathbf{\Phi}), \tag{42}$$

$$\dot{\mathbf{A}}_2 = -\frac{g^2}{m}(\mathbf{A}_1\Phi^2 - \mathbf{\Phi}\mathbf{A}_1 \cdot \mathbf{\Phi}). \tag{43}$$

The equation of motion for the Higgs field is

$$\ddot{\mathbf{\Phi}} = -g^2[(\mathbf{A}_1^2 + \mathbf{A}_2^2)\mathbf{\Phi} - (\mathbf{A}_1 \cdot \mathbf{\Phi}\mathbf{A}_1 + \mathbf{A}_2 \cdot \mathbf{\Phi}\mathbf{A}_2)] - \frac{1}{2}\frac{\partial V}{\partial \mathbf{\Phi}} \cdot \tag{44}$$

From the above we can check that $(\mathbf{A}_1^2 + \mathbf{A}_2^2)$ is a constant of the motion. The canonical momenta are

$$\mathbf{p}_{A_1} = m\mathbf{A}_2, \tag{45}$$

$$\mathbf{p}_{A_2} = -m\mathbf{A}_1, \tag{46}$$

and

$$\mathbf{p}_\phi = 2\dot{\Phi}. \tag{47}$$

The Hamiltonian is given by

$$H = \frac{p_\phi^2}{4} + g^2\left[\left[\mathbf{A}_1^2 + \frac{p_{A_1}^2}{m^2}\right]\Phi^2 - (\mathbf{A}_1 \cdot \Phi)^2 - \frac{(\mathbf{p}_{A_1} \cdot \Phi)^2}{m^2}\right] + V(\phi). \tag{48}$$

Hamilton's equations are

$$\dot{\mathbf{A}}_1 = \frac{g^2}{m^2}(\mathbf{p}_{A_1}\Phi^2 - \Phi\mathbf{p}_{A_1} \cdot \Phi), \tag{49}$$

$$\dot{\mathbf{p}}_{A_1} = -g^2(\mathbf{A}_1\Phi^2 - \Phi\mathbf{A}_1 \cdot \Phi), \tag{50}$$

$$\dot{\Phi} = \frac{\mathbf{p}_\phi}{2}, \tag{51}$$

$$\dot{\mathbf{p}}_\phi = -2g^2[(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi - (\mathbf{A}_1 \cdot \Phi\mathbf{A}_1 + \mathbf{A}_2 \cdot \Phi\mathbf{A}_2)] - \frac{\partial V}{\partial \Phi}. \tag{52}$$

As $m\mathbf{A}_2$ is the canonical momentum corresponding to $\mathbf{A}_1$ and $-m\mathbf{A}_1$ the canonical momentum corresponding to $\mathbf{A}_2$, the system has only six degrees of freedom. It appears that H, $(\mathbf{A}_1^2 + \mathbf{A}_2^2)$, and the three constants of the motion corresponding to the Gauss' law constraint are the only constants of the motion, for NACSH.

However, for the system to be integrable, there must be one more constant of the motion. At this stage it is not clear if such a constant exists. We therefore carry out the Painlevé test for integrability, choosing the scalar Higgs potential to be a quartic of the form $V(\phi) = (\lambda/4)(\phi^2 - v^2)^2$.

It is clear that each of the three components of the gauge fields $\mathbf{A}_1$ and $\mathbf{A}_2$ obeys a first-order differential equation and that each of the three components of the Higgs field obeys a second-order differential equation. So we require twelve arbitrary constants in the local representation of the general solution, if NACSH were integrable.

Carrying out the Painlevé analysis, we first determine the leading behaviour by setting

$$\mathbf{A}_1 = \mathbf{a}_1(t - t_0)^{\alpha_1}, \tag{53}$$

$$\mathbf{A}_2 = \mathbf{a}_2(t - t_0)^{\alpha_2}, \tag{54}$$

$$\mathbf{\Phi} = \mathbf{a}_\phi(t - t_0)^{\alpha_\phi}. \tag{55}$$

We find that the leading behaviour is given by

$$\mathbf{A}_1 = \beta_1\mathbf{a}, \tag{56}$$

$$\mathbf{A}_2 = \beta_2\mathbf{a}, \tag{57}$$

$$\mathbf{\Phi} = \mathbf{a}(t - t_0)^{-1}. \tag{58}$$

with $\mathrm{a}^2 = -4/\lambda$.

For the resonance analysis we consider the next-to-leading order singularity by setting

$$\mathbf{A}_1 = \beta_1\mathbf{a} + \xi_r(t - t_0)^r, \tag{59}$$

$$\mathbf{A}_2 = \beta_2\mathbf{a} + \zeta_r(t - t_0)^r, \tag{60}$$

$$\mathbf{\Phi} = \mathbf{a}(t - t_0)^{-1} + \eta_r(t - t_0)^{r-1}. \tag{61}$$

and

$$\mathbf{\Phi} = \mathbf{a}(t - t_0)^{-1}[1 + q(t - t_0)^{r-1}]. \tag{64}$$

From the general expansion about the leading behaviour, we find that

$$\mathbf{A}_1 = \beta_1 \mathbf{a}, \tag{65}$$

$$\mathbf{A}_2 = \beta_2 \mathbf{a}, \tag{66}$$

$$\mathbf{\Phi} = \mathbf{a}(t - t_0)^{-1}[1 + \sum_i q_i (t - t_0)^i], \tag{67}$$

where $\mathbf{a}^2$, and all the q_i's except q_4, are determined. The arbitrary constants are $t_0, \beta_1, \beta_2, q_4$ and the direction of $\mathbf{a}$. Thus we have six arbitrary constants, while twelve are needed for integrability. Hence the Painlevé test indicates that NACSH is chaotic.

3.2 *The Yang–Mills–Chern–Simons–Higgs system*

The Lagrangian for YMCSH is given by

$$L = -\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu a} + \frac{m}{2} \epsilon^{\mu\nu\alpha} \left[F^a_{\mu\nu} A^a_\alpha - \frac{g}{3} \epsilon_{abc} A^a_\mu A^b_\nu A^c_\alpha \right] + D_\mu \phi^\dagger_a D^\mu \phi_a - V(|\Phi|). \tag{68}$$

Corresponding to this Lagrangian the equations of motion are

$$D_\mu F^{\mu\nu a} + m\epsilon^{\nu\alpha\beta} F^a_{\alpha\beta} = ig[D^\nu \phi^\dagger T_a \phi - \phi^\dagger T_a D^\nu \phi], \tag{69}$$

$$D_\mu D^\nu \phi^a = -\frac{\partial V}{\partial \phi^*_a}. \tag{70}$$

The $\nu = 0$ component gives the following Gauss' law constraint

$$\tfrac{1}{2}(\mathbf{A}_1 \times \dot{\mathbf{A}}_1 + \mathbf{A}_2 \times \dot{\mathbf{A}}_2 + 2m\mathbf{A}_1 \times \mathbf{A}_2) = -\mathbf{\Phi} \times \dot{\mathbf{\Phi}}. \tag{71}$$

Once again YMCSH, the DS in this case, is obtained by considering spatially homogeneous fields. Further, setting $\mathbf{A}_0 = 0$, the equations of motion corresponding to the gauge fields become

$$\ddot{\mathbf{A}}_1 + 2m\dot{\mathbf{A}}_2 + 2g^2(\mathbf{A}_1 \Phi^2 - \mathbf{\Phi}\mathbf{A}_1 \cdot \mathbf{\Phi}) + g^2(\mathbf{A}_1 \mathbf{A}_2 \cdot \mathbf{A}_2 - \mathbf{A}_2 \mathbf{A}_1 \cdot \mathbf{A}_2) = 0, \tag{72}$$

$$\ddot{\mathbf{A}}_2 + 2m\dot{\mathbf{A}}_1 + 2g^2(\mathbf{A}_2 \Phi^2 - \mathbf{\Phi}\mathbf{A}_2 \cdot \mathbf{\Phi}) + g^2(\mathbf{A}_2 \mathbf{A}_1 \cdot \mathbf{A}_1 - \mathbf{A}_1 \mathbf{A}_1 \cdot \mathbf{A}_2) = 0. \tag{73}$$

One constant of the motion for YMCSH is seen to be $\mathbf{A}_2 \cdot \dot{\mathbf{A}}_1 - \mathbf{A}_1 \cdot \dot{\mathbf{A}}_2 + m(\mathbf{A}_1{}^2 + \mathbf{A}_2{}^2)$.

The three-component Higgs field obeys the following equation of motion:

$$\ddot{\Phi} = -g^2[(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi - (\mathbf{A}_1 \cdot \Phi \mathbf{A}_1 + \mathbf{A}_2 \cdot \Phi \mathbf{A}_2)] - \frac{1}{2}\frac{\partial V}{\partial \Phi}. \tag{74}$$

We therefore see that there are nine second-order differential equations for YMCSH, one for each of the three components of $\mathbf{A}_1, \mathbf{A}_2$ and Φ. This means that if the system were integrable, there must exist 18 arbitrary constants in the Painlevé analysis, which may be carried out along lines identical to that in the case of NACSH. It turns out that YMCSH does not pass the Painlevé test for integrability, either.

4. Numerical studies on the non-abelian-Chern–Simons–Higgs system

In contrast to the abelian case, we have seen that the non-abelian counterpart is not integrable, even in the absence of the YM term. To support this fact numerically [20] we present in figures 2(a)–(d), phase plots corresponding to NACSH. The calculations were performed using a Runge–Kutta fourth order algorithm with adaptive step size. In the case of NACSH, since $\mathbf{A}_1$ and $\mathbf{A}_2$ are conjugate to each other, we need to consider only $\mathbf{A}_1$, $\mathbf{A}_2$, Φ and its canonical conjugate $\mathbf{p}_\phi$. Renaming these fields for convenience as x, y, Φ and g respectively, we display in figures 2(a) and (b), plots of ϕ_1 vs g_1 respectively with initial conditions $\mathbf{x} = \mathbf{y} = (0, 1, 1)$ and $\Phi = \mathbf{g} = (1, 0, 1)$. This choice of initial conditions automatically satisfies the Gauss' law constraint. The parameter values

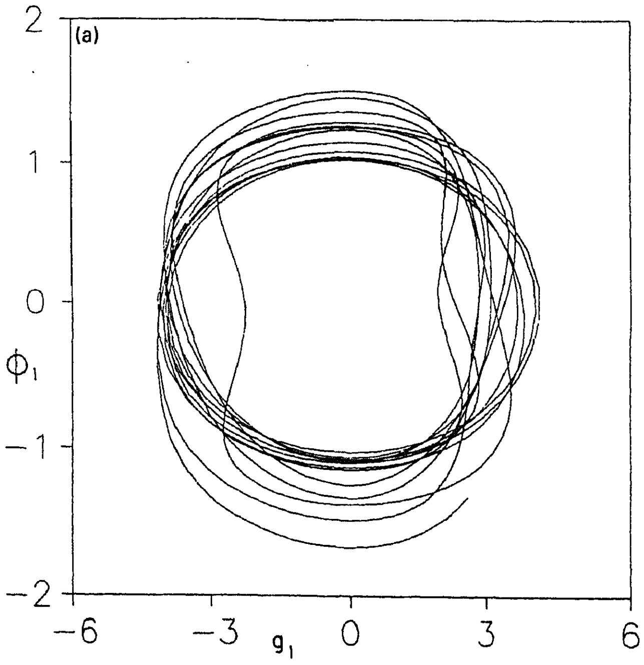


Figure 2(a). Plot of ϕ_1 vs g_1 for the NACSH system with parameters $\lambda = \nu^2 = m = g^2 = 1$ and the initial conditions $\mathbf{x} = \mathbf{y} = (0, 1, 1)$ and $\Phi = \mathbf{g} = (1, 0, 1)$.

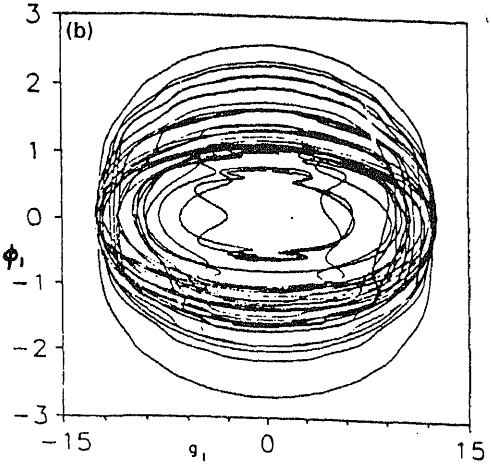


Figure 2(b). Plot of ϕ_1 vs g_1 for the NACSH system with parameters $\lambda = \nu^2 = m = 1$ and $g^2 = 10$ and the initial conditions $\mathbf{x} = \mathbf{y} = (0, 1, 1)$ and $\Phi = \mathbf{g} = (1, 0, 1)$.

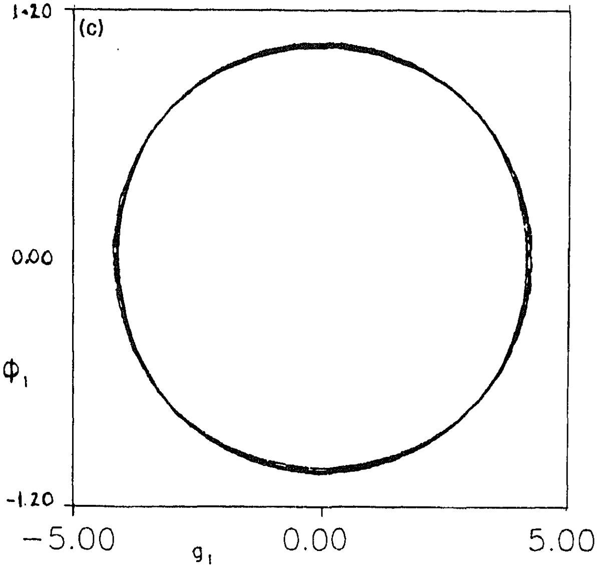


Figure 2(c). Plot of ϕ_1 vs g_1 for the NACSH system with parameters $\lambda = \nu^2 = g^2 = 1$ and $m = 10$ and the initial conditions $\mathbf{x} = \mathbf{y} = (0, 1, 1)$ and $\Phi = \mathbf{g} = (1, 0, 1)$.

esponding to these plots are $\lambda = \nu^2 = m = 1$. However, in figure 2(a) $g^2 = 1$ and in) $g^2 = 10$. Clearly, while the general pattern is the same in both cases, the plots get ser as g^2 increases. In figure 2(c), we display ϕ_1 vs g_1 for the same initial conditions parameter values except that $g^2 = 1$ and $m = 10$. From figures 2(a), (b) and (c) it is r that as m increases some "regularity" in the behaviour of the phase plot is brought

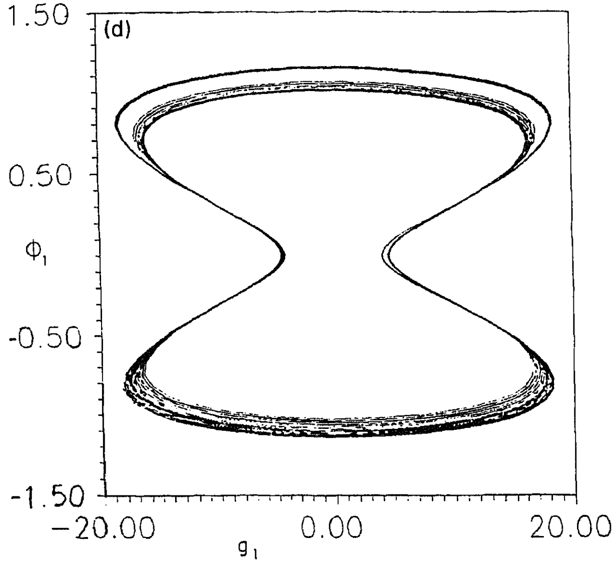


Figure 2(d). Plot of ϕ_1 vs g_1 for the NACSH system with parameters $\nu^2 = m = g^2 = 1$ and $\lambda = 500$ and the initial conditions $\mathbf{x} = \mathbf{y} = (0, 1, 1)$ and $\Phi = \mathbf{g} = (1, 0, 1)$.

about (in the sense that it is not so space-filling). Thus, for large m the system is close to being integrable. In figure 2(d) the plot of ϕ_1 vs g_1 is displayed with all initial conditions and parameter values the same as in figure 2(a) except that $\lambda = 500$. The figure suggests that as λ increases there is a tendency towards regularity.

We now turn our attention to YMCSH. Without displaying the phase plots we comment that as in the NACSH case, an increase in g^2 makes the phase plots more space-filling. This is to be expected as YM is anyway a chaotic system. While an increase in λ brings about more 'regularity' in the Higgs sector, it seems to have no significant effect on the gauge sector.

We now report on results obtained in the non-abelian case by means of an extensive study [20] of the Lyapunov exponents. We first consider the NACSH case. In order to compare our results with those in [18], we must scale the NACSH variables in (56–58), as follows:

$$\mathbf{A}_1 \to g\mathbf{A}_1, \tag{75}$$

$$\mathbf{A}_2 \to g\mathbf{A}_2, \tag{76}$$

$$\Phi \to \frac{g}{\sqrt{m}}\Phi. \tag{77}$$

This reduces the number of parameters from three to one, and correspondingly the equations of motion to

$$\dot{\mathbf{A}}_1 = [\mathbf{A}_2\Phi^2 - \Phi(\mathbf{A}_2\cdot\Phi)], \tag{78}$$

$$\dot{\mathbf{A}}_2 = -[\mathbf{A}_1\Phi^2 - \Phi(\mathbf{A}_1\cdot\Phi)], \tag{79}$$

$$\ddot{\Phi} = -[(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi - (\mathbf{A}_1\cdot\Phi\mathbf{A}_1 + \mathbf{A}_2\cdot\Phi\mathbf{A}_2)] - \frac{\kappa}{2}\Phi(\Phi^2 - \nu'^2) \tag{80}$$

with $\kappa = \lambda m/g^2$. Further, without loss of generality, we set the scaled variable v' to be unity. The scaled Lagrangian becomes

$$L = (\dot{\mathbf{A}}_1\cdot\mathbf{A}_2 - \dot{\mathbf{A}}_2\cdot\mathbf{A}_1) + \dot{\Phi}^2 - [(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi^2 - (\mathbf{A}_1\cdot\Phi)^2 - (\mathbf{A}_2\cdot\Phi)^2] - \frac{\kappa}{4}(\Phi^2 - 1)^2. \tag{81}$$

The corresponding energy function is given by

$$E = \dot{\Phi}^2 + [(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi^2 - (\mathbf{A}_1\cdot\Phi)^2 - (\mathbf{A}_2\cdot\Phi)^2] + \frac{\kappa}{4}(\Phi^2 - 1)^2. \tag{82}$$

In the YMCSH situation, using the same scaling as above we have the following equations of motion:

$$\frac{1}{m}\ddot{\mathbf{A}}_1 + 2\dot{\mathbf{A}}_2 + 2(\mathbf{A}_1\Phi^2 - \Phi\mathbf{A}_1\cdot\Phi) + \frac{1}{m}(\mathbf{A}_1\mathbf{A}_2\cdot\mathbf{A}_2 - \mathbf{A}_2\mathbf{A}_1\cdot\mathbf{A}_2) = 0, \tag{83}$$

$$\frac{1}{m}\ddot{\mathbf{A}}_2 - 2\dot{\mathbf{A}}_1 + 2(\mathbf{A}_2\Phi^2 - \Phi\mathbf{A}_2\cdot\Phi) + \frac{1}{m}(\mathbf{A}_2\mathbf{A}_1\cdot\mathbf{A}_1 - \mathbf{A}_1\mathbf{A}_1\cdot\mathbf{A}_2) = 0 \tag{84}$$

and

$$\ddot{\Phi} = -[(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi - (\mathbf{A}_1\cdot\Phi\mathbf{A}_1 + \mathbf{A}_2\cdot\Phi\mathbf{A}_2)] - \frac{\kappa}{2}\Phi(\Phi^2 - 1). \tag{85}$$

In contrast to NACSH there still remain two parameters here, namely, κ and m. The corresponding scaled Lagrangian is

$$L = \frac{1}{2m}(\dot{\mathbf{A}}_1^2 + \dot{\mathbf{A}}_2^2) + (\dot{\mathbf{A}}_1\cdot\mathbf{A}_2 - \dot{\mathbf{A}}_2\cdot\mathbf{A}_1) + \dot{\Phi}^2 - \frac{1}{2m}[\mathbf{A}_1^2\mathbf{A}_2^2 - (\mathbf{A}_1\cdot\mathbf{A}_2)^2] + [(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi^2 - (\mathbf{A}_1\cdot\Phi)^2 - (\mathbf{A}_2\cdot\Phi)^2] - \frac{\kappa}{4}(\Phi^2 - 1)^2 \tag{86}$$

while the energy function is given by

$$E = \frac{1}{2m}(\dot{\mathbf{A}}_1^2 + \dot{\mathbf{A}}_2^2) + \dot{\Phi}^2 + \frac{1}{2m}[\mathbf{A}_1^2\mathbf{A}_1^2 - (\mathbf{A}_1\cdot\mathbf{A}_2)^2] + [(\mathbf{A}_1^2 + \mathbf{A}_2^2)\Phi^2 - (\mathbf{A}_1\cdot\Phi)^2 - (\mathbf{A}_2\cdot\Phi)^2] + \frac{\kappa}{4}(\Phi^2 - 1)^2. \tag{87}$$

We have examined the variation of the maximal Lyapunov exponent as the values of energy and κ are changed, for a large range of initial conditions. These conditions were chosen to satisfy the Gauss' law constraint. Choosing a one-variable ansatz of the form

$$\mathbf{A}_1 = \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{A}_2 = \begin{pmatrix} 0 \\ x \\ 0 \end{pmatrix} \quad \Phi = \begin{pmatrix} -x \\ x \\ 0 \end{pmatrix} \quad \dot{\Phi} = \frac{1}{2}\begin{pmatrix} x \\ x \\ 0 \end{pmatrix} \tag{88}$$

for NACSH and supplementing them with

$$\dot{\mathbf{A}}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \dot{\mathbf{A}}_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{89}$$

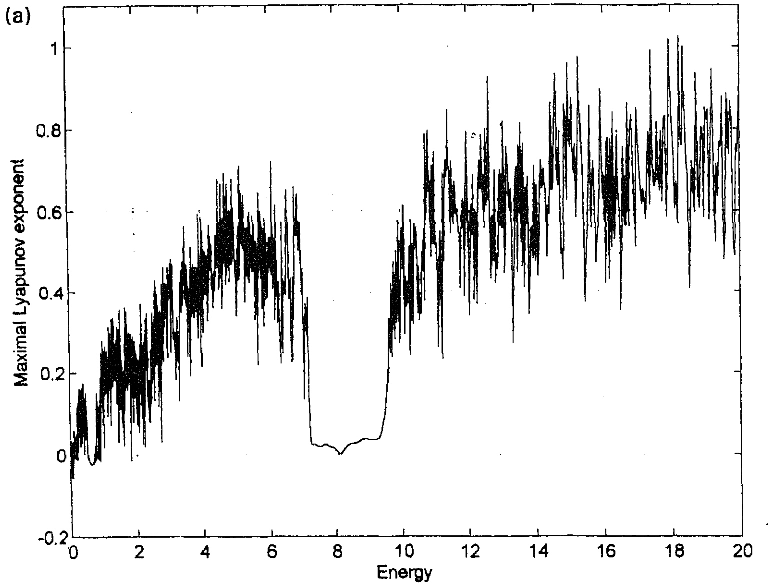


Figure 3(a). Variation of the maximal Lyapunov exponent with energy for NACSH for $\kappa = 0$.

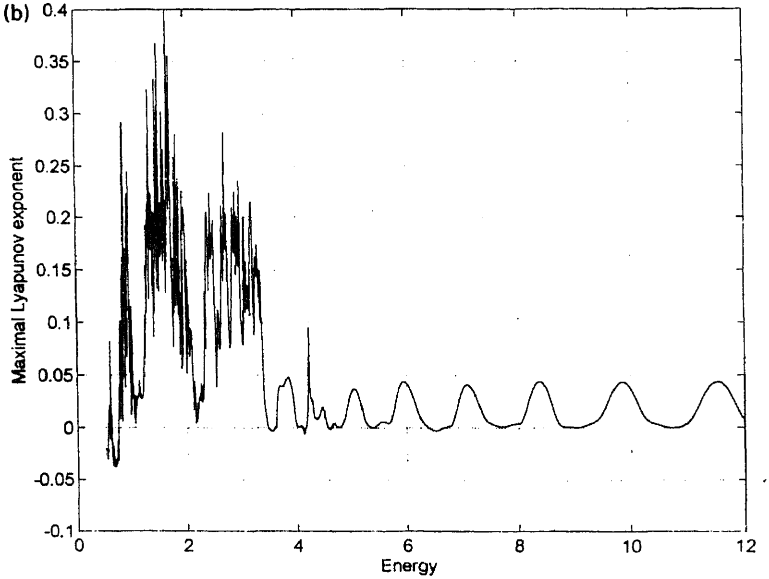


Figure 3(b). Variation of the maximal Lyapunov exponent with energy for NACSH for $\kappa = 5$.

for YMCSH, we have varied x to obtain several initial conditions in the energy region of interest. In figures 3(a) and (b), we show the behaviour of the maximal Lyapunov exponent as a function of energy for $\kappa = 0$ and 5. For $\kappa = 0$, i.e., in the absence of the

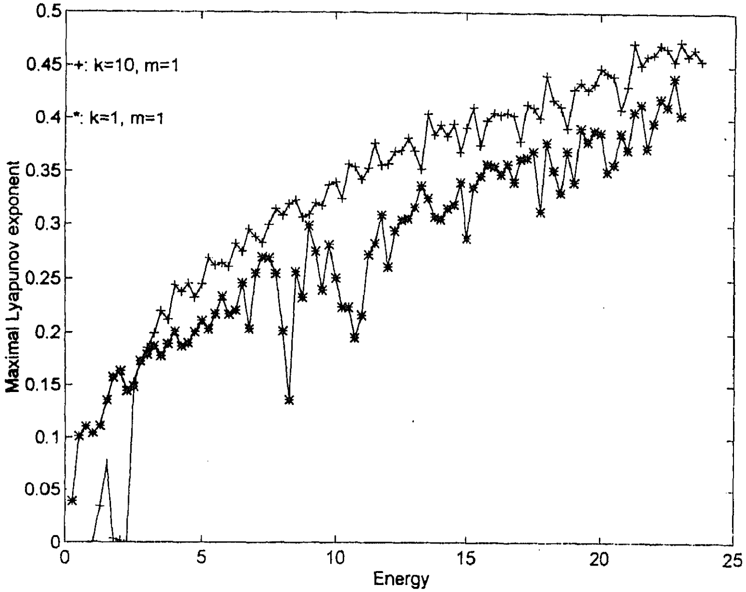


Figure 4. Comparison of maximal Lyapunov exponent vs energy for YMCSH for the one-variable ansatz for $\kappa = 1, m = 1$ (* represents the curve) with $\kappa = 10, m = 1$ (+ represents the curve).

Higgs potential, the system is mostly chaotic as seen from the positive value of the Lyapunov exponent, except for one window of regularity, in the energy region $7 \leq E \leq 9$. While more transitions from order to chaos appear as κ increases, the situation becomes dramatic for large κ as in figure 3(b), where oscillatory order-chaos behaviour appears even from $E = 3$ onwards. This is the effect of an increase in the coefficient m of the topological term. Further, we see that while order-chaos transitions occur on all scales in YMCS [4], this is not true in NACSH. This suggests that the quartic YM coupling is responsible for this behaviour (the occurrence of order-chaos transitions on all scales). A more detailed analysis of the YMCSH system is therefore called for, and we proceed to report on these numerical investigations.

Figure 4 shows the variation of the maximal Lyapunov exponent as a function of energy for $\kappa = 1$ and 10 and with $m = 1$. The graphs reveal that for large κ (be it due to large Higgs coupling λ or small YM coupling g) the system exhibits more regularity for low energies (in the sense of an almost vanishing maximal Lyapunov exponent). The absence of order-chaos transitions on all scales in YMCSH shows that the quartic YM coupling is *not* responsible for this feature observed in the YMCS system. Also, our results are to be contrasted with earlier work [1, 18, 24] in which it is reported that a YMH system becomes more regular as λ increases, regardless of the energy regime investigated.

We have determined the fraction of phase space that is regular by considering [21] about 100 initial conditions for energies ranging from 1 to 10. We find that for small κ, NACSH becomes chaotic more or less monotonically as the energy is increased. For large κ, however, there is no such simple behaviour. For the NACSH system, all the calculations in this regard were carried out with a straightforward Runge–Kutta fourth order routine, preserving constants of the motion to an accuracy of one part in 10^5. However, for calculations of exponents for YMCSH an adaptive step-size Runge–Kutta routine was necessary to ensure energy conservation to the same degree of accuracy as in the case of NACSH.

5. Discussion and conclusions

From the numerical studies reported here on the abelian and non-abelian Chern–Simons–Higgs systems, we arrive at some very interesting conclusions relating to the rich variety of phase space behaviour in these systems, some of these being counter-intuitive. The boundedness of the motion in phase space seems to be a common feature for both the YMH and NACSH systems. This is in contrast to the situation in the abelian Higgs system. This feature is thus independent of the details of the non-abelian nature of the couplings, i.e., whether it is of the YM type or the CS type.

Contrary to what we may expect intuitively, it is not true that an increase in κ "regularizes" the gauge term at all energies. If the increase in κ is due to one in the Higgs coupling, the degree of regularity may be intuitively expected to increase; but if the increase in κ is due to a decrease in the gauge coupling, it is not established that further regular islands appear for all small (non-zero) values of the gauge coupling. As borne out by our studies, however, an increase in the scaled parameter κ enhances the regularity only for small energies [21]. This feature can be understood better if we realise that appearance of regularity also depends on the nature and availability of the phase space. When more regions in phase space become accessible, it is plausible that randomness could increase and the KAM tori structure could get further broken.

Another aspect is that as the energy increases, the maximal Lyapunov exponent increases in magnitude in the YMCSH system. Thus the YM term takes over for large energies and the CS term produces the "oscillatory" effect between regularity and chaos.

While large windows of regularity exist in NACSH, this is not true for YMCSH. This could perhaps be traced back to the interplay between the "oscillatory effect" of the CS term, the "regularizing effect" of the Higgs term, and the completely chaotic signature of the YM term.

The message from our investigations is as follows. In contrast to systems with a few degrees of freedom, relatively complex systems of the sort we have considered, governed by non-trivial symmetries, can display unusual and rich dynamical behaviour, thanks to the intricate interplay between the various coupling constants. As even CSH (the simplest DS that we have considered) shows, its integrable nature is lost on including a linear coupling to the Maxwell term (to yield MCSH). As the number of degrees of freedom increases, the intricacies and the chaotic nature of phase plots increases, as may be expected. The next step would be the study of the case of a full-fledged field theory, where it would be necessary to identify good quantifiers of chaos. At the present stage our understanding of the connections (if any) between the analytical tests of integrability like the ARS conjecture and numerical methods as in the Fermi–Pasta–Ulam approach is rather poor. Further, several problems relating to the significance of chaos in field theories quantized using semi-classical quantization prescriptions remain open. The role of chaos in field theories that support topological solutions is not understood completely, although some interesting results have been obtained (see, for example, [11, 25]); and finally, the implications of the lack of integrability in quantum field theories are yet to be examined.

Acknowledgments

The authors are very grateful to S Seshadri and E W B Dias for help in the preparation of the manuscript and V Balakrishnan for a critical reading of the manuscript. One of the authors (CM) was supported by the University Grants Commission under the Research Scientist scheme, and (BB) thanks the University Grants Commission for partial support.

References

[1] S G Matinyan, G K Savvidi and N G T Savvidi, *Z. Eksp. Teor. Fiz.* **80**, 830 (1981)
[2] A A Belavin and V E Zakharov, *JETP Lett.* **25**, 567 (1977)
[3] M Jimbo, M D Kruskal and T Miwa, *Phys. Lett.* **A92**, 59 (1982)
R S Ward, *Phys. Lett.* **A102**, 279 (1984)
[4] A Giansanti and P D Simic, *Phys. Rev.* **D38**, 1352 (1988)
[5] E S Nikolaevskii and L N Schur, *JETP Lett.* **36**, 218 (1982)
[6] B V Chirikov and D L Shepelyanskii, *Pis'ma Zh. Eksp. Teor. Fiz.* **34**, 171 (1981)
[7] M J Ablowitz, A Ramani and H Segur, *Lett. Nuovo. Cimento.* **23**, 333 (1978)
[8] G K Savvidy, *Nucl. Phys.* **B246**, 302 (1984)
[9] R Marcinek, E Pollak and J Zakrzewski, *Phys. Lett.* **B327**, 67 (1994)
[10] S G Matinyan, E B Prokhorenko and G K Savvidi, *JETP Lett.* **44**, 138 (1986)
[11] T Kawabe and S Ohta, *Phys. Rev.* **D41**, 1983 (1990)
[12] E Fermi, J R Pasta and S Ulam, *Los Alamos Sci. Lab Tech.* Report No. LA-1940 (1955) (unpublished)
[13] B Dey, C N Kumar and A Sen, *Int. J. Mod. Phys.* **A8**, 1755 (1993)
[14] C N Kumar and A Khare, *J. Phys.* **A22**, L849 (1989)
[15] M Toda, *Phys. Lett.* **A8**, 335 (1974)
P Brumer and W Duff, *J. Chem. Phys.* **65**, 3566 (1976)
[16] T Kawabe and S Ohta, *Phys. Rev.* **D44**, 1274 (1991)
[17] T Kawabe, *J. Phys.* **A26**, L1131 (1993)
[18] T Kawabe, *Phys. Lett.* **B343**, 254 (1995)
[19] B A Bambah, S Lakshmibala, C Mukku and M S Sriram, *Phys. Rev.* **D47**, 4677 (1993)
[20] M S Sriram, C Mukku, S Lakshmibala and B A Bambah, *Phys. Rev.* **D49**, 4246 (1994)
S Lakshmibala, in *Computational aspects in chaos and nonlinear dynamics* edited by G Ambika and V M Nandakumaran (Wiley Eastern, New Delhi, 1994) pp 37–47
[21] C Mukku, M S Sriram, J Segar, B A Bambah and S Lakshmibala (to be published in *J. Phys.* **A**)
[22] A Khare, *Fortschr. Phys.* **38**, 507 (1990)
[23] S K Paul and A Khare, *Phys. Lett.* **B191**, 389 (1987)
[24] S G Matinyan, *Sov. J. Part. Nucl.* **16**, 226 (1985)
[25] M P Joy and M Sabir, *Pramana – J. Phys.* **38**, L91 (1992)
J Segar and M S Sriram, *Phys. Rev.* **D53**, 3976 (1996)

Chaos in Josephson junctions

G AMBIKA
Department of Physics, Maharajas College, Cochin 682 011, India

Abstract. A detailed analysis of the control space characterization of phase locked states and chaotic attractors in Josephson junctions is presented, based on a model that includes both quadratic damping and cosine interference terms. In addition, some novel features of the nonlinear characteristics of the junction like evolution of basin boundaries, bifurcation structure analysis and scaling behaviour of Lyapunov exponent are discussed.

Keywords. Josephson junctions; chaos; fractal basin boundary; phase locking; Lyapunov exponent and scaling index.

PACS Nos 05.40; 05.45; 74.40; 85.25

1. Introduction

During the past two decades, Josephson junction (JJ) has served as a powerful paradigm in understanding low dimensional chaos in nonlinear systems. This is mainly because the equation modeling the dynamics of the junction, in its simplest form, represents a minimal system for the observation of chaos and it is isomorphic to a variety of physically interesting situations like the forced pendulum, pinned charge density waves and phase locked loops. Moreover, it is a remarkably fertile field exhibiting a large variety of nonlinear effects ranging from phase locking and sub harmonic generation to chaotic transitions of different kinds like period doubling, crisis etc. Further, the nonlinear behaviour in Josephson junctions is easily accessible for experimental measurements in addition to circuit simulations and numerical analysis. This system is well known to be technologically important due to its possible applications in many electronic devices like parametric amplifiers, local oscillators, mm wave detectors and voltage standards.

Experimental observation of chaos in these systems dates back to 1968 [1]. Historically speaking, JJ oscillators revealed a striking noise rise phenomena that could not be explained in terms of thermal fluctuations. In superamps, an increase in the amplitude of the driving signal led to broad-band voltage fluctuations with an equivalent noise temperature of $\approx 10^4$ K [2]. Huberman and his group suggested that this can be understood in terms of chaotic solutions of the nonlinear junction dynamics [3, 4]. Later the period doubling phenomena was observed above a critical voltage in tunnel junctions [5] and irregularities in the I-V characteristics of the junction were reported as signatures of chaos [6–8]. Experimental results on deterministic chaos in JJ [9] and nonlinear properties of the junction dynamics [10], have been recently reviewed.

Apart from experimental studies, theoretical investigations have been carried out in JJ based on equations that model the dynamical characteristics of the junction. The most popular and well studied of such models is the RSJ model or the Stewart–McCumber model [11] in which the resistance R that accounts for the current of quasi particles and the self capacitance C of the junction are shunted across the device. Then the response of the junction to dc and rf biasing can be represented by an equivalent circuit. In simple models the resistance R is taken to be independent of the junction voltage $v(t)$. Survey of chaos in JJ centered around this model through digital simulations, numerical investigations and circuit realizations of the modeling equations, abound in the literature [12–14]. Such studies could identify regions in the parameter space of the junction where nonlinearity generates the most complex behaviour. In the chaotic regime, a variety of intermittent regions [15] are observed after a complete period doubling sequence. In actual experiments such interesting and complicated structure are usually washed out by thermal noise [7], which induces dynamics very much like chaos called thermally induced chaos [16].

A few studies reported in later years based on some modified models are worth mentioning, the major modifications being in the damping term of the equation. If in a current biased junction, the shunted resistance R is taken to be nonlinear and vary as $1/v(t)$, the damping term in the equation of motion becomes quadratic [17, 18]. Another type of nonlinear damping mechanism arises when the conditions are such that the interference effects between the pair and quasi-particle current should be taken into account. This adds a cosine term to the usual damping term [19, 5]. Studies in the linear RSJ model with the cosine term reveal that its presence sensitively governs the responses of the system [20, 21]. Since chaotic responses of the system arise due to the nonlinear effects of the system, we feel it is relevant to include both these for a complete analysis of the possible modes of the system.

In this paper, we report the results of such a study on a quadratic RSJ model with cosine interference terms. We focus our studies along four main tracks, viz. characterization of phase locked states and chaotic regimes, bifurcation structure studies to bring out the relevance of the dc bias and cosine terms, evolution of the basin boundaries when the system asymptotes to multiple attractors and the scaling behaviour of the order parameter for the transition from periodicity to chaos.

In § 2, we discuss the salient features of the junction treated as a dynamical system in the quadratically damped model with the cosine term. We then proceed to a preliminary search procedure for the transition from periodicity to chaos using Melnikov analysis, to provide lowest threshold for the onset of chaos that can be prescribed using the unperturbed solutions in terms of the relevant parameters of the system. The JJ as a dynamical system turns out to be a system with multiple attractors that occur both in the same well and also in different potential wells. Each attractor will have its own separate basin of attraction and the nature of evolution of the boundary between them forms an interesting study. We discuss the evolution of basin boundaries as the system develops period doublings, crises and tangles in § 4.

Section 5 presents the results of a numerical analysis with a view to isolate the windows of phase locked states in control parameter space are included. The chaotic regions are then marked out according to their 'degree' of chaos that is estimated based on the magnitude of the Lyapunov exponents (LE). The nature of the scenario leading to chaos is studied in the following section by plotting the detailed bifurcation structures with the frequency of the drive as the tuning parameter. The crossover from periodic to chaotic behaviour is signalled by the change of the maximal LE from negative to positive value. Hence it has the status of an order parameter in the transition. The scaling behaviour of the LE is therefore relevant, in general, from a phase transition point of view. We numerically compute the LE near the transition point for different types of scenarios that occur in the present system such as period doubling, intermittency between phase locked states, crisis etc. The results are given in § 7, and compared with numerical work based on other models, and with available experimental data, in the concluding section.

2. Josephson junction as a dynamical system

The Josephson junction is an SIS junction consisting of two superconducting electrodes separated by a tunnelling barrier. The tunnelling of electron pairs produces a dissipationless current I_s, which flows at zero bias voltage and given by [22]

$$I_s(t) = I_0 \sin x(t), \tag{2.1}$$

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = 2ev(t)/\hbar, \tag{2.2}$$

where $x(t)$ is the phase difference between the super conducting order parameters on both sides of the junction. I_0 is the maximum supercurrent that is geometry, temperature and material dependent, $v(t)$ is the voltage across the junction. In addition to this current of Cooper pairs, for nonzero voltages, there is a current of quasi particles modelled as a current through the ohmic resistance R_1 and the interference effects are taken care of by a term proportional to cos $x(t)$ through a resistance R_2. Under a dc and rf drive, the equation of motion can thus be written as [23]

$$C\frac{\mathrm{d}v}{\mathrm{d}t} + \left[\frac{1}{R_1(v)} - \frac{1}{R_2(v)}\cos x\right]v + I_0 \sin x = I_{\mathrm{dc}} + I_1 \sin \Omega t. \tag{2.3}$$

We introduce scaled variables and dimensionless parameters in the following way. t is replaced by $\omega_0 t$, where ω_0 is given by

$$\omega_0^2 = (2eI_0)/(\hbar C). \tag{2.4}$$

The resistance R_1 is taken as a constant in the simple RSJ model. But the general expression for $1/R_1$ as given in [24] shows that its dependence on v is through an exponential function such that as an approximation, we can take

$$R_1(v) = k_1 \bigg/ \left|\frac{\mathrm{d}x}{\mathrm{d}t}\right| \quad \text{and} \quad R_2(v) = k_2 \bigg/ \left|\frac{\mathrm{d}x}{\mathrm{d}t}\right|. \tag{2.5}$$

If we define

$$k = (k_1 C)^{-1}, \quad \varepsilon = k_1/k_2, \quad I = I_{\mathrm{dc}}/I_0, \quad A = I_1/I_0 \quad \text{and} \quad \omega = \Omega/\omega_0, \tag{2.6}$$

equation (2.3) then becomes

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + k\left|\frac{\mathrm{d}x}{\mathrm{d}t}\right|\frac{\mathrm{d}x}{\mathrm{d}t}(1 - \varepsilon \cos x) + \sin x = I + A \sin \omega t. \tag{2.7}$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = \omega. \tag{2.8}$$

The control parameter space of the system (2.8) is 5 dimensional, spanned by $(k, \varepsilon, I, A, \omega)$. At the origin of this parameter space, the system is isomorphic to the classical nonlinear pendulum with the same phase space structure as the pendulum in the $(x_1 - x_2)$ phase plane. By analogy with the pendulum, we characterize the periodic modes of the system by (p, q) so that $p = 0$ corresponds to the oscillatory modes and $p \neq 0$ are rotational modes. q specifies the periodicity of the mode relative to the rf drive, while p gives the number of rotations made in one period. The qualitatively different trajectories of these two modes are separated by the separatrix orbit. The fixed points of the system alternate along the x_1-axis, the stable ones or centres at $(2n\pi, 0)$ and the unstable ones or saddles at $((2n+1)\pi, 0)$, for integer n values. The stable manifold W^{s} and the unstable manifold W^{u} of the separatrix connect the saddles at $(n\pi, 0)$ and $(-n\pi, 0)$. It is easy to see that the above dynamics has an infinite series of potential wells with minima at $(2n\pi, 0)$ and maxima at $((2n+1)\pi, 0)$ as its background [25].

When the control parameters are small enough, (2.8) can be considered as a perturbation of the pendulum structure discussed above. Then W^{u} and W^{s} of the saddles reorganize themselves and may touch each other resulting in a heteroclinic tangency. The threshold value for this to occur can be computed analytically by the Melnikov analysis [26]. The separation between W^{s} and W^{u} is measured by the Melnikov function $M(t_0)$ in terms of the solutions of the unperturbed separatrix, x_{10} and x_{20}. (t_0 here refers to the initial time that fixes the particular Poincare section where the trajectories are visualized.) The vanishing of $M(t_0)$ then provides the required threshold. For the system in (2.8), we have [27]

$$x_{10} = \pm 2 \tan^{-1}(\sinh \tau) \quad \text{and} \quad x_{20} = \pm 2 \operatorname{sech} \tau, \tag{2.9}$$

where $\tau = t - t_0$ and

$$M(t_0) = \int_{-\infty}^{\infty} x_{20}[-kx_{20}^2(1 - \varepsilon \cos x_{10}) + A \sin \omega(t - t_0) + I] \, \mathrm{d}\tau. \tag{2.10}$$

Evaluating the integral using (2.9) we get

$$M(t_0) = k(\pm 2\pi\varepsilon \mp 4\pi) \pm 2\pi I \pm 2\pi A \operatorname{sech}(\pi\omega/2) \sin \omega t_0. \tag{2.11}$$

The threshold value for the parameter A for tangency is

$$A_M = |k(2 - \varepsilon) - I| \cosh(\pi\omega/2). \tag{2.12}$$

As A increases beyond A_M, the W^{s} and W^{u} of the separatrix accumulate on each other due to their transverse intersections, forming a heteroclinic tangle. Thus there are Smale horseshoes in the underlying dynamics, although the existence of horseshoe does not in most cases imply asymptotic chaos. The initial conditions which are near the invariant set of the horseshoe may experience a chaotic transient but will ultimately settle down to motion on an attractive periodic orbit. On the other hand, if the points are eventually fed back into the same region, a chaotic attractor will be formed. Near the tangle therefore the system develops a sensitivity to initial conditions.

The dependence of the threshold on other relevant parameters can be deduced from (2.12). In the absence of the dc bias, the threshold for a given frequency ω, is generally higher and therefore a nonzero value of I should bring down the transition curve for onset of chaos in the $A - \omega$ plane. This has been numerically verified in a previous study [27]. So also, by fixing I at 0.5 and $\varepsilon = 0.8$, the threshold is minimum for $k = 0.4$ corresponding to moderate damping in the system. When ε is not included, however, the minimum occurs for low damping. Similarly for $\varepsilon = 0.8$ and $k = 0.4$, the threshold has its minimum at $I = 0.5$. We, therefore restrict our studies in subsequent sections to parameter values $k = 0.4$, $\varepsilon = 0.8$ and $I = 0.5$, since this is presumably the most favourable choice for onset of chaotic behaviour in the system.

3. Fractal basin boundaries

It is clear from the discussion in the previous section that the system in (2.8) has multiple attractors. The attractors are all simple fixed points when there is no perturbation and the separatrix defines the smooth boundary of their basins. If we concentrate on the central potential well, the oscillatory modes developed under forcing will have bounded basins around $(0, 0)$. The escape from the well normally leads to $p \neq 0$ modes which, with the damping may settle down to $p = 0$ modes in distant wells or cross well oscillations as well as chaotic attractors. The nature and pattern of basin changes in each well should be the same but when the central well is under consideration, the basin of attractors in other wells form part of the escaping basin. Once the threshold for tangle is reached, the smooth boundary starts developing structures and becomes a fractal, even when the asymptotic state is a periodic mode in a distant well. *A priori* knowledge about how attractor basins evolve is of value when we address the question of control mechanisms. So also, the extent of basin erosion due to forcing can tell us how far the system has chances of surviving in a noisy environment and this is relevant to stability related engineering problems [28, 29].

Our previous studies of (2.8) indicates that for small ω values, the stable mode excited is a $(0, 1)$ one which we call S_n. Choosing a window of phase space defined by $0 < x_1 < 8$ and $-3 < x_2 < 3$, we analyse the bounded basin of this mode numerically. With $k = 0.1$, $\omega = 0.3$, $I = 0.0$ and $\varepsilon = 0.0$, we try to see how the basin evolves as A is slowly tuned. For this, the system in (2.8) is integrated numerically using fourth order Runge–Kutta–Gills (RKG) algorithm and checked for the required periodicity. For the central well $(0, 1)$ modes, the x_2 values averaged over a cycle is zero and $-\pi < x_1 < \pi$. The integrations are repeated for a grid of 80×60 initial conditions and the asymptotic state identified after 100 drive cycles using the above information. The initial points are then colour-coded to distinguish the escaping and bounded basins [30].

For the set of basin portraits displayed in figure 1 the bounded basin of S_n is in red while the escaping basin belonging to attractors in other wells is yellow. As we increase A

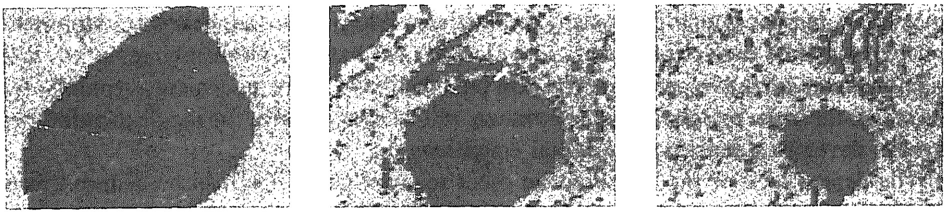


Figure 1. The bounded basin for the nonresonant $(0, 1)$ mode (red) in the central well and the escaping basin (yellow) belonging to attractors in other wells for increasing values of A viz. 0.2, 0.5 and 0.8.

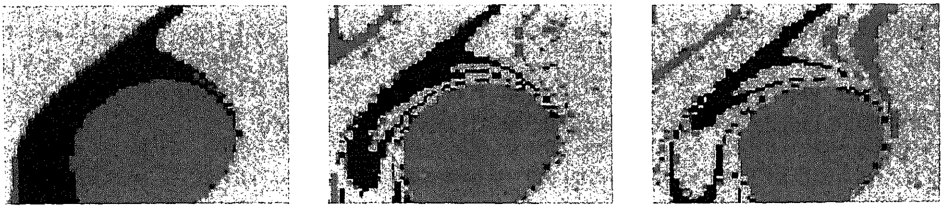


Figure 2. The sharing of the bounded basin by the coexisting S_n and S_r modes. With $\omega = 0.6$ as A is tuned from 0.3 to 0.34, the basin of the latter, indicated by the blue colour, gets reorganized, while that of the former suffers a progressive shrinking.

beyond the threshold A_M, the reorganization of the basin due to the tangle affects only the escaping basin and the nonresonant bounded basin suffers only a progressive shrinking. It is to be noted in this connection that S_n does not undergo period doubling and the onset of chaos is finally due to a reverse boundary crisis [31]. Here the chaotic attractor that exists for $A > 1.017$ collides with the saddle at the basin boundary resulting in the sudden disappearance of the attractor together with its basin [32].

For higher values of ω, a periodic resonant $(0, 1)$ mode S_r is born along with its unstable counterpart via a saddle node bifurcation before the Melnikov tangency occurs. The bounded basin in the central well is now shared by S_n and S_r and we have the phenomenon of resonant hysteresis. In figure 2, blue regions indicate the newly created basin of S_r. Above the threshold, increase of A results in a drastic reorganization of this basin due to the incursion of the escaping basin. However before mass erosion of the basin starts, S_r undergoes period doubling and the resulting chaotic attractor disappears via a boundary crisis [31] that occurs at $A = 0.411$. Beyond this we find only the bounded basin of S_n and the escaping basin shared by the chaotic attractor and attractors in other wells.

However for values of $\omega \approx 1$, S_n has already disappeared and the bounded basin belongs to S_r alone [27]. Then above A_M long finger like structures start penetrating its basin and massive erosion of the whole bounded basin takes place. These events are clear from the series in figure 3. The bounded basin in the central well is coloured yellow while the escaping basin belonging to attractors in the successive wells on the right are given green, violet, red and blue respectively. The brighter shades of the same colours indicate basins of modes in wells on the left. The white regions lead to rotational modes or chaotic attractors. Similar results for erosion of bounded basins have been reported by Soliman and Thompson [32, 33] for periodically driven and linearly damped nonlinear oscillators.

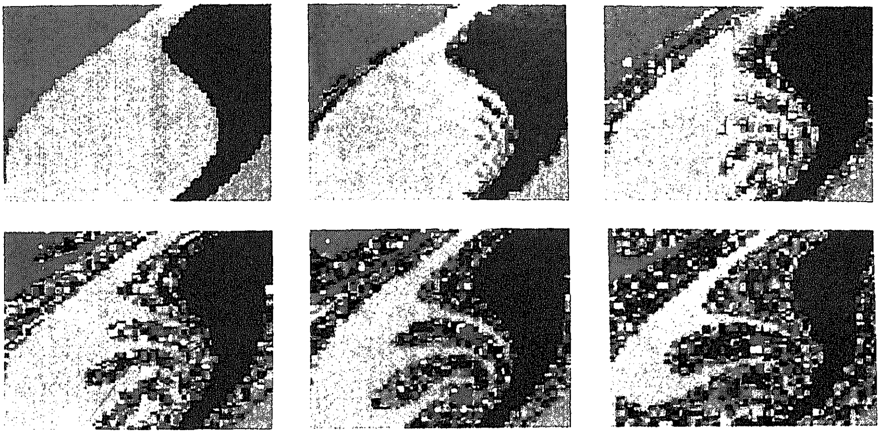


Figure 3. Evolution of the fractal basin of the resonant mode for $\omega = 0.8$. The development of the heteroclinic tangle is coincident with the erosion of the basin due to the penetration of long finger like structures. Here the A values taken are 0.2, 0.35, 0.45, 0.5, 0.55 and 0.6.

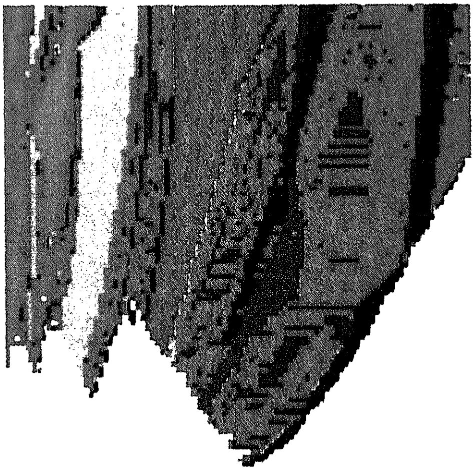


Figure 4. Phase locked regions in the (A, ω) plane. The yellow regions correspond to $(0, 1)$ mode while the red regions lead to chaotic or a periodic states. For the colour code used for other (p, q) modes, see the text.

. Phase locked states and chaotic regimes in the control space

n this section we discuss the occurrence of phase locked states in (2.8), with $k = 0.4$, $= 0.5$ and $\varepsilon = 0.8$. The $A - \omega$ plane is then scanned in the range 0.2–1.2, with a step ize of 0.01 [34]. For each point thus chosen, (2.8) is integrated numerically using fourth rder RKG scheme with a time step equal to one fiftieth of the drive period. The first 000 values are discarded to shake off transients and the remaining checked for phase

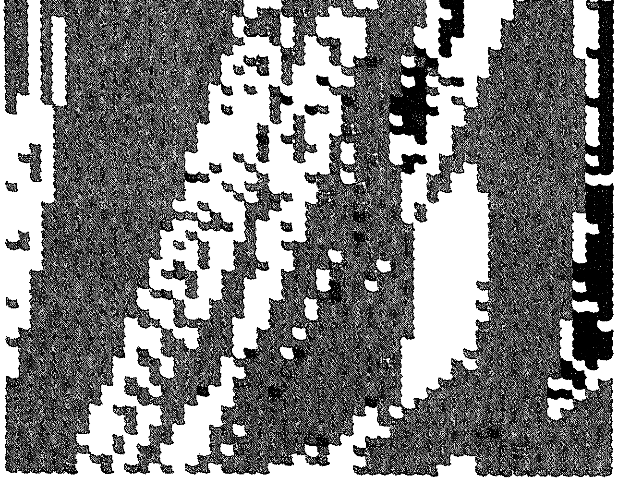


Figure 5. Chaotic domains in the (A, ω) plane. Here the red regions related to periodic states with $\lambda_1 < 0$, while the other shades correspond to $\lambda_1 > 0$; for details of colours used see the text.

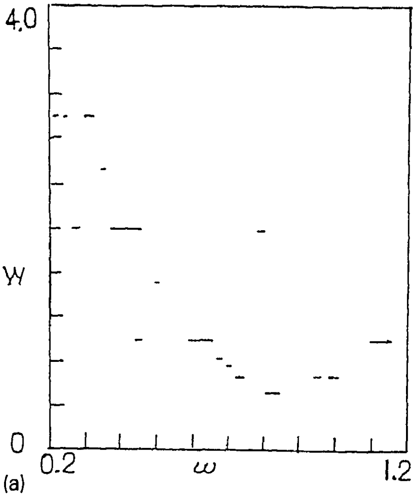


Figure 6a. Devil's staircase of the phase locked modes, with the winding number $W = p/q$ plotted against ω. Here $A = 0.76$, $k = 0.4$, $\varepsilon = 0.8$ and $I = 0.5$.

locked states (p, q) based on the equations

$$x_1(t + 2\pi q/\omega) = x_1(t) + 2\pi p,$$
$$x_2(t + 2\pi q/\omega) = x_2(t). \tag{4.1}$$

The regions of different (p, q) values in the (A, ω) plane are shown in figure 4. Yellow

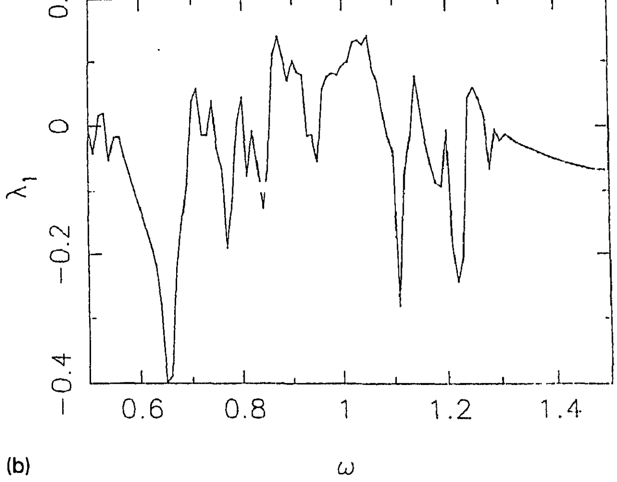


Figure 6b. Variation of the maximal Lyapunov exponent as a function of the rf drive frequency, with $A = 1.0$ and other parameter values as in 6a.

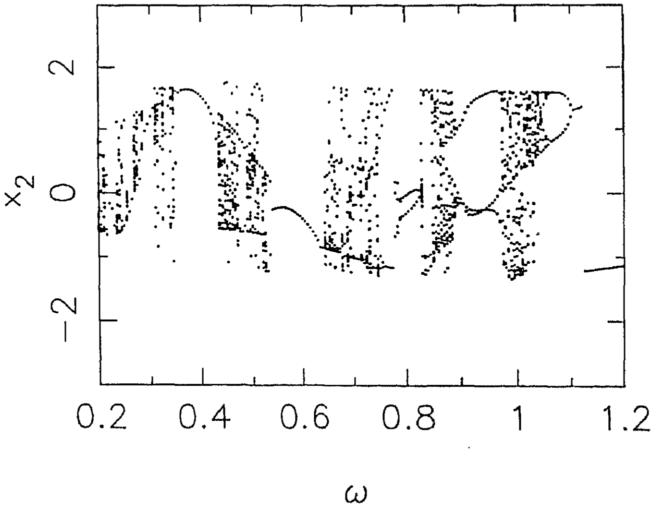


Figure 7. Bifurcation diagram with ω as the tuning parameter for $A = 0.765$, $k = 0.4$, $I = 0.5$ and $\varepsilon = 0.8$.

regions correspond to parameter values where the $(0, 1)$ mode is stable, while the red regions lead to periodic or chaotic states. The regions where the period-doubled state of the $(0, 1)$ mode viz $(0, 2)$ stabilizes are painted dark blue. The $(1, 1)$ and $(1, 2)$ rotational modes are stable for (A, ω) values chosen from the light blue and green regions respectively. Similarly the $(2, 1)$ and its corresponding 2-cycle $(2, 2)$ stabilize for values from the orange and blue magenta regions respectively. The magenta color corresponds to the stable $(3, 1)$ mode. Although computations were done for $q \leq 32$ and $p \leq 5$, modes with $p > 3$ and $q > 4$ have very narrow regions in the parameter plane and hence are not clearly seen in the figure.

The regions in the (A, ω) plane that are marked red in figure 4 where phase locking has not been observed can be further characterized by computing the Lyapunov exponent [35, 36]. We restrict the (A, ω) plane further to regions of mostly chaos i.e. A and ω values in the range 0.5–1.0 (figure 5). The 'strength' of chaos in each region is marked out by giving different shades based on the value of λ_1. The red regions in the figure correspond to negative values of λ_1 and therefore belong to periodic regimes. The major chaotic windows are the dark blue regions where $0.1 < \lambda_1 < 1.0$ and the yellow ones with $0.01 < \lambda_1 < 0.1$. There are also some narrow and isolated spots of cyan for which $0.001 < \lambda_1 < 0.01$ and blue cyan for which λ_1 lies between 10^{-4} and 10^{-3}.

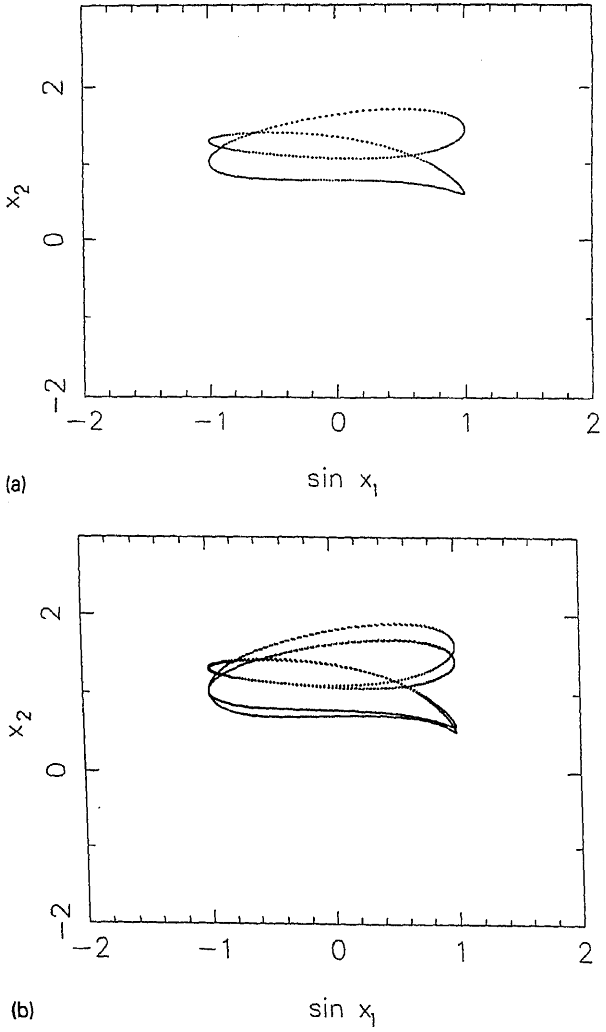


Figure 8a, b. (*Continued*)

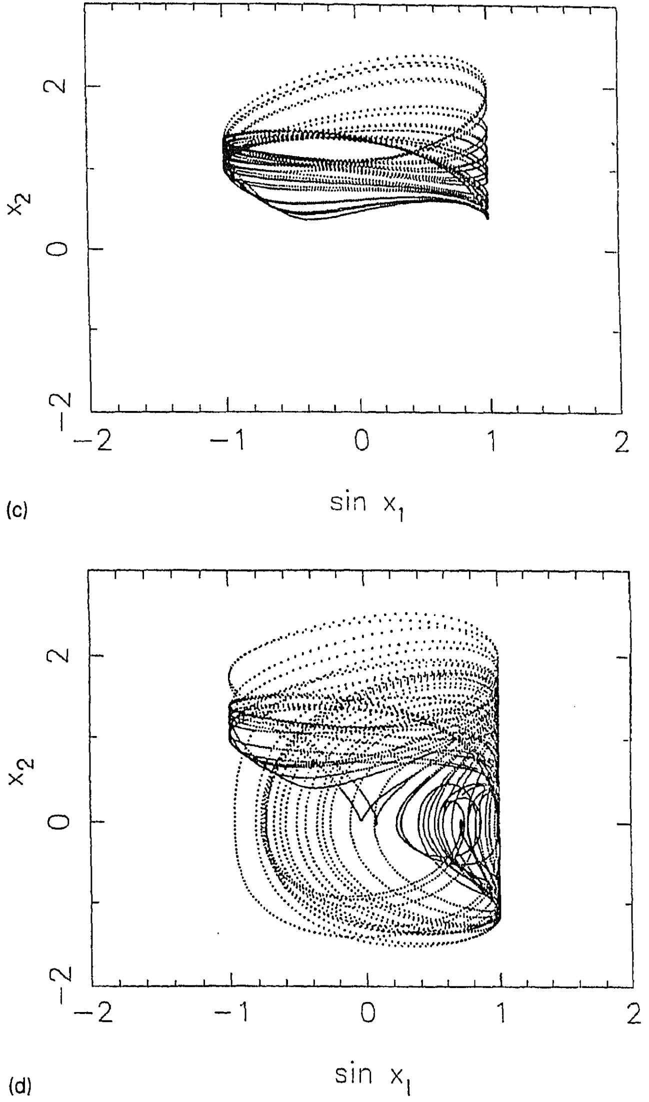


Figure 8. Phase portraits in the sin $x_1 - x_2$ plane, showing the period doubling route to chaos followed by interior crisis resulting in the expansion of the attractor. In **(a)** $A = 0.765$ and $\omega = 1.07$, the periodic $(2, 1)$ mode is seen as **(b)** ω is tuned to 1.06, $(2, 1)$ period doubles to $(2, 2)$, **(c)** the chaotic attractor of $(2, 1)$ for $\omega = 1.04$ and **(d)** the attractor after the interior crisis for $\omega = 1.0$.

The phase-locked states that correspond to the nonzero voltage steps of the junction, ve relevance in voltage standards and can be further characterized by calculating the nding number of the modes, which is defined as

$$W = \langle x_2 \rangle / \omega. \tag{4.2}$$

A plot of W against ω is given in figure 6a. It is clear that the system gets locked into a single step for a range of values of ω, forming a devil's staircase structure. Here the major steps shown in the figure follow the Farey tree sequence but there are gaps in between steps corresponding to random behaviour due to shuttling between two steps. The variation of λ_1 with ω for $A = 1.0$, is given in figure 6b. The phase locked windows and the chaotic regimes can be made out clearly.

5. Bifurcation studies

The different types of scenarios that the system follows in its passage to chaos can be pursued further by plotting the bifurcation diagrams. It is clear from § 3 that this is decided basically by the frequency ω. Hence in this section we analyse the bifurcations that the system undergoes as ω is slowly tuned from above. The last thousand values of the integrator sampled at the frequency of the drive are plotted against ω to get the bifurcation diagram given in figure 7.

Near $\omega = 1$, the resonant $(0, 1)$ mode that is excited would undergo a period doubling as ω is decreased. However it is difficult to capture the successive period doublings till chaos is reached. This is mainly because the basin of attraction of the $(0, q)$ mode very soon becomes narrow and eaten away by the basins of other (p, q) modes which coexist with $(0, q)$ modes. These $p \neq 0$ modes also undergo the period doubling bifurcations and reach chaos. As ω is tuned further the chaotic attractor thus produced expand and merge together due to an interior crisis [32]. This sequence of events can be made out clearly from the set of phase portraits of the attractors in the $\sin x_1 - x_2$ plots included in figure 8. Here with $A = 0.765$ and $\omega = 1.07$, the system settles down to a $(2, 1)$ mode. The first period doubling occurs at $\omega = 1.06$ followed by another at 1.056 etc. The accumulation point is at 1.04 followed by interior crisis and the extended attractor after this for $\omega = 1.00$ is also shown in the figure.

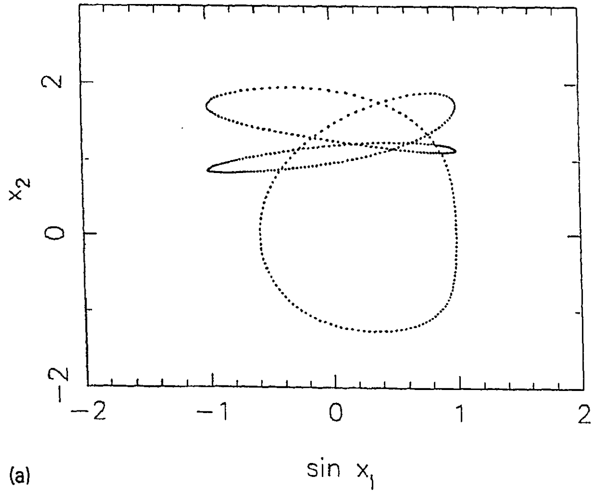


Figure 9a. *(Continued)*

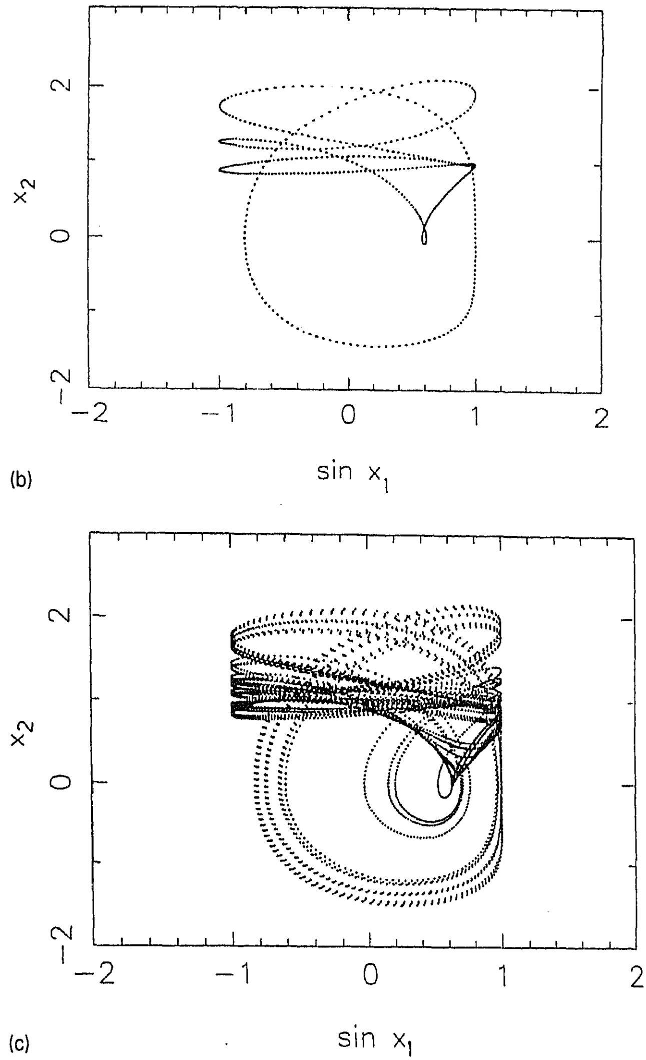


Figure 9. The chaos between two stable windows of phase locked states. In (**a**) with $A = 1.1$, $\omega = 0.85$, (2, 1) mode is stable while in (**b**) for $\omega = 0.796$, (3, 1) mode stabilizes. For $\omega = 0.80$ (**c**), the chaotic attractor due to random shuttling between the above two modes is seen.

In addition to the above two types of chaotic attractor, the system supports a third type which arises from the random shuttling between two phase locked states. This happens every time one phase locked state becomes unstable and before another one is created. For $A = 1.1$, periodic states are observed for ω values 0.85 and 0.796. In between for $\omega = 0.80$, the shuttling type of behaviour is seen. This is illustrated by the portraits in figure 9. The same pattern of events is seen repeated in several regions of parameter space. However for very small ω values, only the nonresonant (0, 1) mode is supported and this disappears via a reverse boundary crisis mentioned earlier.

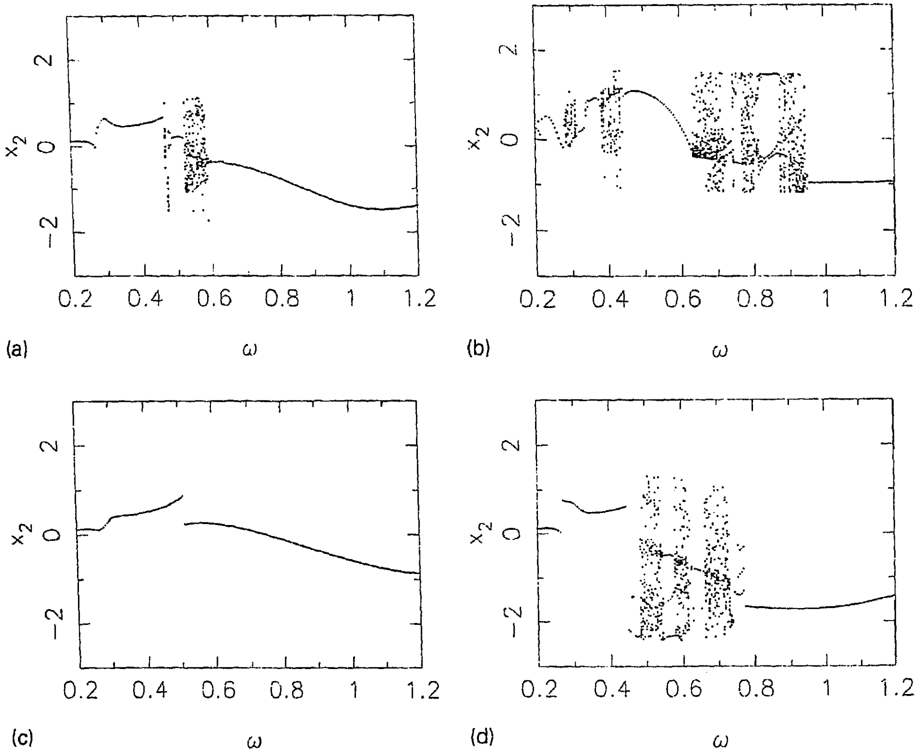


Figure 10. Bifurcation diagrams as in figure 7, but with (**a**) $I = 0.0$ (**b**) $\varepsilon = 0.0$ (**c**) $I = 0.0$, $\varepsilon = 0.0$ and (**d**) $I = 0.0$, $\varepsilon = 0.0$ but $k = 0.1$.

When the above analysis is repeated for $I = 0.0$, it is found that most of the chaotic regions are wiped out and the (0, 1) mode remains stable up to $\omega = 0.6$. The effect is less spectacular when ε is put equal to zero, resulting in the broadening of a few phase locked windows. When both I and ε are excluded the periodic modes predominate and chaotic regions reappear only when k is reduced to values ≈ 0.1 (figure 10). These events are supported by the analysis based on Melnikov threshold discussed earlier.

6. Scaling index for the Lyapunov exponent

Since λ_1 passes through zero at the transition point ω_c of any scenario, its scaling behaviour near this point is studied in detail here. If we write

$$\lambda_1 \approx |\omega - \omega_c|^{\nu} \tag{6.1}$$

then ν is called the scaling index. Computationally the first step to obtain ν is to identify the transition point accurately. This is easy for the period doubling scenario, since by tuning ω in very fine steps the value for which λ_1 is sufficiently small can be located. A log-log plot of λ_1 and $|\omega - \omega_c|$ furnishes the index ν as its average slope (figure 11). Our studies in this direction indicates that the value of ν for most cases is 1.0 [37],

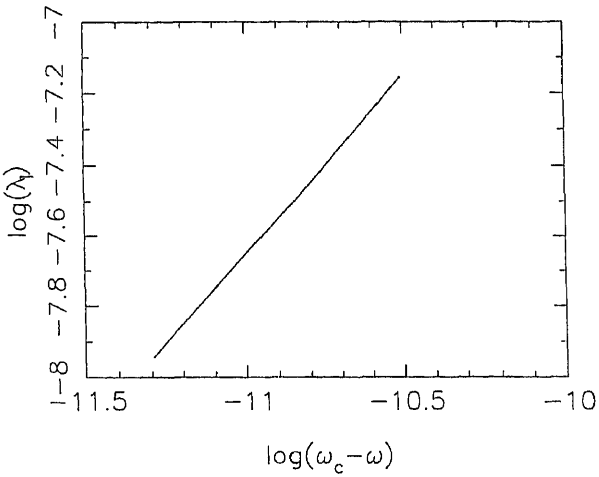


Figure 11. Scaling behaviour of Lyapunov exponent. Here $A = 0.765$ and $\omega_c = 1.054510642311$; the average slope is $\nu = 1.00011$.

Table 1. Scaling indices for the LE values during the transitions to chaos via the different mechanisms mentioned in the text. All the entries except the fifth are for the period doubling routes while the fifth one is for the transitional type of chaos. For the last entry alone, I and ε are taken to be zero with $k = 0.1$.

A	ω	ω_c	ν
0.5	0.913519734–0.913519743	0.913519743424	0.9999
0.9	1.071781040–1.071781090	1.07178113667069	0.9998
0.9	0.889031880–0.889031960	0.889031960230	0.9966
0.765	1.054510642280–1.054510642306	1.054510643111	1.0001
1.1	0.79679753021092–0.7967975302110	0.796797530210909	1.0001
0.765	0.545662031341–0.545662031346	0.5456620313487151	1.0001

establishing linear scaling for the maximal Lyapunov exponent in the system. The onset of the transitional type of chaos due to shuttling between two phase locked states is also found to follow a linear scaling. The results for a few typical cases are condensed in table 1.

We also scan regions of crisis along the ω-line, where the λ_1 value is not zero at ω_c. The critical value of interior crisis is located by changing ω values in very small steps near the crisis value or by looking at the actual trajectories or Poincare maps until an expansion of the attractor is observed. After the crisis the λ_1 values fluctuates about the crisis value. The ω_c value for reverse boundary crisis is similarly masked by the presence of long-lived chaotic transients.

7. Conclusion

In this paper, the results of a detailed analysis of the phase locked states and chaotic regimes in the quadratically damped model of Josephson junctions is presented. The major results reported are the following.

periodicity to chaos indicates that dc bias can lower the threshold as a whole. So also, low values of the damping factor k favours the transition. This is supported by the bifurcation studies presented. Near the threshold, the boundary of the bounded basin develops fractal nature and for high values of ω long finger like structures start penetrating the basin leading to its rapid erosion. The chaotic regions are characterized further based on the magnitude of the maximum Lyapunov exponent. The transition to chaos for the cases studied in this work, is associated with a linear scaling relation for this exponent.

Direct experimental measurements of chaos in actual junctions do not unfortunately match with the parameters chosen for numerical work. However, there are a few exceptions. Using Pb–Te–Pb junctions with a Te thickness of 40–75 nm exposed to 35 GHz radiation, the occurrence of chaos as signalled by a rise in the equivalent noise temperature has been reported [8]. In this case, the parameters are $k = 0.37$, $\omega = 0.18$ and I is tuned from 5–7 mA. They also observed intermittency between phase locked states which leads to noise rise at the end of the voltage steps as ω was scanned between 0.16 and 0.14. It is also interesting to note that with no dc bias, no rise in noise was observed at $\omega = 0.2$. Our results are more or less in confirmity with this, since with $I = 0.0$, we find most of the chaotic regions are wiped out (figure 10).

The devil's staircase picture of the phase locked states, the scaling behaviour of the LE across the transition point and the characterization of the chaotic solutions in the parameter space are some of the novel features in the present work. However parameter space characterization of phase locked states have been reported in earlier models also [9, 22]. We would like to mention that in [9], the $I - 1/k$ plane and in [22], the $I - \varepsilon$ plane with $A = 0.61$, $\omega = 0.8$ and $k = 0.2$ are chosen. They report regions of periodic modes, intermittent chaos connecting two metastable states called transfer crisis, regions of narrow band chaos from period doubling and wide band chaos where the x_2 values are not confined. The last case, we feel, must correspond to chaos after interior crisis.

Our studies relate to the (A, ω) plane where A and ω are external parameters that can be tuned easily unlike k and ε which are internal parameters characterizing a particular junction. The (A, ω) plane is chosen in [38], with the linear RSJ model, for $k = 0.2$ and $I = 0.0$. The threshold is lowest for $\omega = 1.0$ and chaos occurs only above $A = 1.0$ for small values of ω. Moreover, oscillatory modes of the $(0, 1)$, $(0, 2)$ and $(0, 3)$ types are seen above the threshold. This is different from our results since we find only windows of rotational modes in the chaotic region above the threshold. So also, the lowest for the threshold corresponds to $\omega = 0.65$–0.7 and near $\omega = 1.0$, the resonant $(0, 1)$ mode is stable. This is because eventhough escape from the central well occurs, the system settles down to the $(0, 1)$ mode in a different well after long chaotic transients. Similar result is seen in Huberman's work [15]. Here the terminology is slightly different since the phase locked states are specified by $n(= p/q)$ rather than by p and q separately. Giving allowance for this, we find the transition curve in [15] is higher than in our work. This, as is clear from the latter part of our work, must be due to the inclusion of dc bias. We get more phase locked windows in regions which are indicated as intermittency in [15]. It seems as if the quadratic damping together with the cosine term increases the tendency of the system to get locked into different frequencies. This is especially so for low values of ω, where broad windows corresponding to $(1, 1)$, $(2, 1)$, $(3, 1)$ etc. are seen in our plot.

It is interesting to note that the type of quadratic damping present in the system, has relevance in a broader sense. Firstly such systems can be reduced to a hamiltonian format [39] and hence its quantum effects can be studied using approximation techniques [40]. Secondly, rapid motion of bodies in fluids of low viscosity gives rise to vortices, whose production involves resistance forces proportional to the square of the velocity. This effect is then called turbulent damping [41]. In general, quadratic damping terms prop up whenever forces of the type kv^2 and $-kv^2$ act on a system, alternatively damping and exciting it for successive half cycles of oscillations.

The study of the nonlinear properties of JJ carries lot of promise both from the point of view of applications and continued research on fundamental problems. Recently intense research is being concentrated on long JJ which is an extension of simple junctions, where a spatial variation of the pair phase is included. This system is known to support soliton propagation and indicates an even richer dynamics [42] due to the inter play between time and space and possible analogies to turbulent systems [43].

In conclusion, the problems we have discussed so far in JJ are defined on the basis of BCS superconductivity. Future work involving new type of high T_c superconductors may bring out completely different nonlinear phenomena due to anisotropy and complicated flux dynamics. Although a few isolated attempts in this direction have already been recorded [44] it must be exciting to plunge into such promising areas in JJ systems.

Acknowledgements

Major part of the computational aspects of the work presented here, has been carried out in the DRDO computer centre, Cochin Univ. of Sci. and Tech., Cochin under a CSIR associateship and at IUCAA, Pune as an Associate. Both are gratefully acknowledged.

References

[1] Mc Cumber, *J. Appl. Phys.* **39**, 3113 (1968)

[2] R Y Chiao, M J Feldman, D W Pedersen, B A Tucker and M T Levinsen, in *Future trends in superconductive electronics*, *AIP Conf. Proc.* **44**, 259 (1978)

[3] B A Huberman, J P Crutchfield and N H Packard, *Appl. Phys. Lett.* **37**, 750 (1980)

[4] M R Beasley and B A Huberman, *Comments Solid State Phys.* **10**, 155 (1982)

[5] N F Pedersen, M R Samuelsen and K Saermark, *J. Appl. Phys.* **44**, 5120 (1973)

[6] K Okuyama, H J Hartfuss and K H Gundlach, *J. Low Temp. Phys.* **44**, 283 (1981)

[7] D C Cronemeyer, C C Chi, A Davidson and N F Pedersen, *Phys. Rev.* **B31**, 2667 (1985)

[8] M Octavio and C R Nasser, *Phys. Rev.* **B30**, 1586 (1984)

[9] M Octavio, *Physica* **A163**, 248 (1990)

[10] N F Pedersen, *Physica* **D68**, 27 (1993)

[11] K K Likharev, *Dynamics of Josephson junctions and circuits* (New York: Gordon Breach, 1986)

[12] R L Kautz and R Monaco, *J. Appl. Phys.* **57**, 875 (1985)

[13] M Octavio, *Phys. Rev.* **B29**, 1231 (1984)

[14] E G Gwinn and R M Westervelt, *Phys. Rev. Lett.* **54**, 1613 (1985)

[15] D D Humieres, M R Beasley, B A Huberman and A Libcaber, *Phys. Rev.* **A26**, 3483 (1982)

[16] A Davidson, B Dueholm and M R Beasley, *Phys. Rev.* **B33**, 5127 (1986)

[17] M Bartucelli, P L Christiansen, N F Pedersen and M P Soerensen, *Phys. Rev.* **B33**, 4686 (1986)

essentially opposite point of view is taken [2, 3] and it is advocated that semi-hard collision processes between the partons (momentum transfer $P_T > 1$ GeV) will dominate the parton thermalization. More precisely, Shuryak [2] has made use of lowest order perturbative QCD matrix elements to arrive at the conclusion that at RHIC energies, gluons would thermalize in a time $\tau_{\mathrm{g}} \approx 0.5\,\mathrm{fm/c}$ whereas the quarks would take $\tau_{\mathrm{q}} \approx 2\,\mathrm{fm/c}$ with the initial temperatures estimated as 500 MeV for gluons and 200 MeV for quarks.

An approach complementary to that involving parton 'collisions' is the one giving thermalization due to 'collective' non-perturbative effects in a quark gluon plasma. Unfortunately, there is no quantum field-theoretic approach to study time-dependent non-perturbative phenomena in a many-body QCD system; hence most of the work has been carried out in a classical framework. Thus, Müller and Trayanov [4] have carried out a classical lattice simulation of pure $SU(2)$ Yang–Mills theory and demonstrated that the system is chaotic. They further argue from the largest Lyapunov exponent that an estimate of the thermalization time is $\tau_{\mathrm{g}} \approx 0.5\,\mathrm{fm/c}$, which is comparable to Shuryak's value. Bhatt [5], Bhatt *et al* [6, 7a] and more recently Sengupta *et al* [8] have respectively used parton hydrodynamic equations and parton particle simulations to demonstrate chaos in non-abelian longitudinal oscillations in a quark gluon plasma; as shown below, the estimate of collective thermalization time from these studies is $\tau \approx 1\,\mathrm{fm/c}$.

It is important to emphasize however, that the thermalization due to semi-hard collisional processes and that due to collective (chaotic) behaviour correspond to physically different parameter domains. The crucial parameter is the plasma parameter $\Lambda_{\mathrm{D}} = n\lambda_{\mathrm{D}}^3$ i.e., the number of particles in a Debye sphere. It is well known that collective behaviour dominates when $\Lambda_{\mathrm{D}} \gg 1$ and collisional effects will be more important when $\Lambda_{\mathrm{D}} \lesssim 1$.

In this review we give a very brief qualitative account of the parton thermalization work done by earlier workers (§ 2) and then go on to describe in some detail the work done by our group on parton thermalization by collective effects (§ § 3–5). Our work is based on the classical parton equations of motion due to Wong [9]. These are introduced in § 3, with a detailed discussion of the model. In § 4, we present the results of our studies based on colour hydrodynamic equations [6, 7, 9] derivable from Wong's model. Section 5 describes the results of our recent particle simulation of Wong's equations and summarises our conclusions regarding thermalization by collective non-abelian effects. Finally, § 6 provides a discussion of the results.

2. Review of earlier work on thermalization

Detailed microscopic models of relativistic heavy-ion collisions, based on QCD, are being extensively used to investigate issues of space-time evolution and subsequent thermalization of collision products. Such models use either quantum mechanical or classical approach to describe the collision process. Models using quantum mechanics are further sub-divided into two groups, viz., (a) string models and (b) Parton–Cascade models. In the string models, [10, 11], the non-perturbative aspect of QCD interaction is described phenomenologically based on the data of soft hadron-hadron collisions (typically a reasonable description when the four momentum transfer P_T is small compared to the QCD scale parameter Λ_{QCD}). In the parton cascade models [3], on the other hand, one uses perturbative QCD which is known to be a satisfactory model of semi-hard and hard ($P_T > \Lambda_{\mathrm{QCD}}$) hadron-hadron interactions. The classical models [4] provide us a complementary picture to the quantum mechanical models since they permit us to incorporate time-dependent non-perturbative aspects of the QCD interactions. In these models, the well known chaotic behaviour of the classical non-abelian gauge theories is linked with the thermalization of the plasma.

Each of the models mentioned above has its own merits and deficiencies. We now briefly review them.

2.1 *String models*

As indicated above, this approach is based on a phenomenological model of soft ($P_T < \Lambda_{\mathrm{QCD}}$) hadron-hadron collisions. In these models [10, 11], the two colliding hadrons are assumed to interpenetrate each other and form a colour flux tube between them. The colour tube then disintegrates into quark-antiquark pairs and gluons in a time scale of order 1 fm/c. The entropy production and thermalization time for the heavy-ion collision problem is then determined by such soft hadron-hadron collision processes. At higher beam energies involving four-momentum transfer $P_T \geq \Lambda_{\mathrm{QCD}}$, the application of such models becomes questionable primarily because the string density increases significantly and the strings begin to overlap. Since the models do not consider the interaction between the strings [10, 11], they begin to lose their validity. At such energies, hard collisions between the hadrons play a very important role and a prediction of the collisional dynamics solely based on soft collision processes may become inaccurate.

There also exists a model [12, 13] offering a similar picture of the nuclear collision at a 'macroscopic' level. This is the condenser plate model of ultra relativistic heavy ion collisions. The picture of the collision offered by such a model may be valid even at $P_T \gtrsim \Lambda_{\mathrm{QCD}}$. In this model, one imagines that the Lorentz-contracted nuclei at ultra relativistic energies, assume the shapes of pancakes with a thickness ~ 1 fm [14]. The nuclei then become essentially transparent and pass through each other because the time to fly past their Lorentz contracted diameters is shorter or comparable to the strong interaction time-scale. The central region of the collision could thus be baryon free. However, the soft gluon exchange process can create colour charge on the receding nuclei producing a colour electric field in the central region. This colour electric field is unstable to the production of quark-antiquark pairs and gluons [15, 16]. Thermalization time of the quark-gluon plasma produced in this way can be estimated using kinetic theory with a parton production rate term evaluated using the Schwinger mechanism. The particle energy density, entropy production rate and equilibration time can be computed using the kinetic theory approach [12, 17, 18]. It is found that the thermalization time is of the order of a few tens of parton production rate time units.

[18] G Ambika, K Babu Joseph and V M Nandakumar, *Mod. Phys. Lett.* **B5**, 519 (1991); **B5**, 1037 (1991)
G Ambika and K Babu Joseph, in *Symmetries and singularity structures* edited by M Lakshmanan and M Daniel (Heidelberg: Springer, 1990)
[19] Y Yao, *Phys. Lett.* **A118**, 59 (1986)
[20] F Palmero and F R Romero, *Phys. Lett.* **A160**, 553 (1991)
F Palmero and F R Romero, *Phys. Lett.* **A179**, 337 (1993)
[21] K Sakai and Y Yamaguchi, *Phys. Rev.* **B30**, 1219 (1984)
[22] R E Harris, *Phys. Rev.* **B10**, 84 (1974)
[23] N F Pedersen and K Saermark, *Physica* **69**, 572 (1973)
[24] M Morisne and A Kanasugi, in *Bifurcation and chaos* edited by A Awrejcewicz (Heidelberg: Springer, 1995) p. 133
[25] G Ambika, *Indian J. Phys.* **A67**, 551 (1993)
[26] E D Rio, A R Lozano and M G Velarde, *Chaos, solitons and fractals* **4**, 255 (1994)
Z J Jing, *SIAM J. Appl. Math.* **49**, 1749 (1989)
W Xiao and X Yao, *Acta. Phys. Sin.* **38**, 137 (1989)
[27] G Ambika, *J. Phys. Condens. Matter* **4**, 4829 (1992)
[28] J M T Thompson and Y Ueda, *Dynamics and stability of systems* **4**, 285 (1989)
[29] M S Soliman and J M T Thompson, *Int. J. Bif. Chaos* **2**, 81 (1992)
[30] G Ambika, *Curr. Sci.* **65**, 870 (1993)
[31] C Grebogi, E Ott and J A Yorke, *Physica* **D7**, 181 (1983)
[32] A N Lansbury and J M T Thompson, *Phys. Lett.* **A150**, 355 (1990)
[33] M S Soliman and J M T Thompson, *Phys. Rev.* **A45**, 3425 (1992)
[34] G Ambika, *Phys. Rev.* **B** communicated
[35] A Wolf, T R Swift, H I Swinney and J A Vastano, *Physica* **D16**, 285 (1985)
[36] T S Parker and L O Chua, *Practical numerical algorithms for chaotic systems* (New York, Springer, 1989)
[37] G Ambika, *Phys. Lett.* **B221**, 323 (1996)
[38] N F Pedersen, in *Superconducting electronics*, NATO ASI series vol. 59 edited by H Weinstock and M Nisenoff (Berlin: Springer, 1989) p. 209
[39] G Ambika, *Phys. Lett.* **A150**, 273 (1990)
[40] G Ambika and V M Nandakumaran, *Phys. Lett.* **A192**, 331 (1994)
[41] P Srinivasan, *Nonlinear mechanical vibrations* (New Delhi: New Age Int. P. Ltd., 1995)
[42] L E Guerrerro and M Octavio, *Phys. Rev.* **B37**, 3641 (1988)
S Rajasekar and M Lakshmanan, *Phys. Lett.* **A147**, 264 (1990)
G Rotoli and G Filatrella, *Phys. Lett.* **A156**, 211 (1991)
[43] F Helsot, B Castaing and A Libchaber, *Phys. Rev.* **B36**, 5870 (1987)
[44] D Browne and B Horovitz, *Phys. Rev. Lett.* **61**, 1259 (1988)

—journal of
physics

February 1997
pp. 655–674

Collective thermalization in quark gluon plasma

S SENGUPTA, J R BHATT*, P K KAW and J C PARIKH*
Institute for Plasma Research, Bhat, Gandhinagar 382 424, India
*Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India

Abstract. A very important question in ultrarelativistic heavy ion collisions is that of thermalization of the high energy density quark gluon plasma formed in the central rapidity region. Different approaches have been adopted by various authors to study this thermalization problem. These include phenomenological string and capacitor plate models, perturbative QCD based parton cascade models and the classical non-perturbative approach. In this paper we briefly review the earlier studies and discuss our work which emphasizes the role of non-perturbative collective effects (classical chaos) in the thermalization of the plasma. In particular, using classical equations of motion of a coloured parton in self-consistent colour fields, we have carried out a $1 + 1$ dimensional simulation of coloured partonic matter. We find that in certain parameter domains, the system exhibits chaotic behaviour in non-abelian plasma oscillations, which then leads to thermalization of the plasma.

Keywords. Quark gluon plasma; collective effects; thermalization; chaos.

PACS No. 12.38

1. Introduction

There is a strong theoretical expectation that collisions of nuclei at ultra relativistic energies would produce a deconfined state of coloured partons called quark gluon plasma (QGP). The fireball (plasma) that is produced during the collision, will evolve in space and time from a pre-equilibrium phase to a thermally equilibrated phase (atleast locally) and will eventually undergo a phase transition to hadronic matter. An important element in this scenario is the process(es) by which the highly excited, dense coloured particle system comes to thermal equilibrium and the time it takes for reaching equilibrium. The study of thermalization is crucial for several reasons. In a fundamental sense, it throws light on the dynamics of the pre-equilibrium phase. Further from a practical viewpoint, once the plasma is in local thermodynamic equilibrium, its space time evolution can be examined using the well-known hydrodynamic approach of either Landau or Bjorken. From the experimental point of view, thermalization has important consequences for signals of quark gluon plasma [1].

A number of studies using quite different dynamic processes have been carried out to understand the thermalization problem. This already indicates that there is a great deal of uncertainty concerning the dynamics in the very early stages of the evolution of the plasma. In one of the approaches extrapolating from *pp* and *pA* studies, soft QCD processes (involving, for example, the dual string model) have been assumed to be

2.2 *Perturbative QCD based models*

When the four momentum transfer $P_T \geq \Lambda_{\text{QCD}}$, methods of the perturbative QCD (pQCD) can be used to describe the hadron-hadron interaction. At collider energies $\sqrt{s} \geq 100\,\text{GeV}\,A$, major component of the pp and $p\bar{p}$ collision spectra can be reproduced by taking semi-hard or hard partonic interaction into account. Therefore, it is reasonable to expect that in the heavy-ion collisions at RHIC or LHC energies hard or semi-hard partonic scattering would play a significant role in describing the evolution of the collisions.

In this approach, each nucleon is initially in a colour coherent state and regarded as a cloud of partons. Each parton carries some fraction of the nucleon's energy and is capable of producing a virtual cascade of partons. Due to hard or semi-hard collisions with another particle, the parton in the virtual cascade can become real and the colour coherence in the nucleon may get broken. Thus, we have the following picture of the heavy ion-collisions: Two fluxes of partons are counter streaming against each other leaving a highly excited dense partonic system in the central region, which is not in thermal equilbrium. Thermalization is established due to interactions among the partons. The phase-space evolution of the partonic system is described by Boltzmann type transport equation for various species of quarks, antiquarks and gluons. The form of collision terms for the different processes are obtained using the perturbative QCD cross-sections.

The initial parton distribution contains mostly gluons and they thermalize faster than the quarks as the gluon-gluon interaction dominates over quark-gluon or quark-quark interaction. It is found from such studies that momentum equilibration of the initially produced parton distribution occurs very early i.e., in a time of the order of the inverse of collision frequency [19]. However, as the parton production continues even after the momentum isotropy is established the chemical equilibration can be established only at a later stage of the collision. The production rates of the partons are calculable by perturbative QCD in the early stages of the collision. The entropy produced per secondary parton as a function of time can be studied within such models and it reaches a plateau for a wide range of energies in a time less than 1 fm/c. Typical estimate for the gluon equilibration time is $\tau_{\text{g}} \sim 0.5\,\text{fm/c}$ and production and equilibration time for quarks is $\tau_{\text{q}} \sim 2\,\text{fm/c}$. Thus, one may expect that the thermalization in heavy-ion collision proceeds through two stages: first gluons get thermalized at a higher temperature ($\simeq 500\,\text{MeV}$) and then quarks thermalize at a lower temperature ($\approx 200\,\text{MeV}$). Some of the predictions of this model can be tested by probing the signals coming from the early phase dynamics of nuclear collisions [20].

2.3 *Classical collective models*

As mentioned above, there are no quantum field theoretic methods available to study time-dependent non-perturbative aspects of collective (many-body) phenomena in QCD. Hence, most efforts in this direction are based on a classical framework. It is well known that a classical theory of elementary particles and fields has ultraviolet problems because of which it is not adequate to describe QCD interactions. However, it is found that at high temperature and in the long-wavelength limit, classical and perturbative QCD treatments give the same results for quark gluon plasma dispersion relations [11]. Therefore, one tends to believe that even at the non-perturbative level, qualitative features of the classical approach will remain valid in the long wave-length limit. Unlike the quantum mechanical models, where non-perturbative effects are included phenomenologically (strings, flux tubes, etc.), in the classical approach the non-perturbative aspects come from the QCD Lagrangian. As mentioned earlier, the thermalization of the plasma is then related to the well known tendency of classical non-abelian gauge fields to become chaotic. Following the earlier work of Matinyan *et al* [21] on development of chaos in classical Yang–Mills theories (in the infrared limit $k \to 0$), recently Müller and Trayanov [4] and Gong [22] have carried out numerical studies of the classical Yang–Mills theories on a lattice in the presence of spatial inhomogeneities ($k \neq 0$). They show that randomly chosen field configurations of $SU(2)$ or $SU(3)$ gauge theories have universal largest Lyapunov exponent which scales with the average energy of the system. It was shown that for an extremly large class of field configurations the universal exponent is given by $\lambda_0 \sim (1/10)g^2\langle E_\mathrm{p}\rangle$ (for $SU(3)$ theories). The entropy for such a system grows linearly with time and is given by the relation

$$S(t) = S_0 + \Sigma_i \lambda_i t. \tag{1}$$

From this relation one can obtain the entropy growth rate $\mathrm{d}S/\mathrm{d}t \geq \lambda_0$. The typical thermalization time found from the entropy production rate is in conformity with the perturbative QCD based expectation.

As indicated in the introduction, there is another classical approach to the problem of equilibration which has been pursued by our group [7, 8]. This approach is motivated by the desire to put quarks and gluons on the same footing in a description of the collective phenomena, in contrast to the above calculations where the quarks are totally ignored. In the rest of this paper we shall present a detailed review of this work.

3. Classical colour parton equations

We now describe the approach followed by our group to investigate collective non-abelian aspects of thermalization in a quark gluon plasma. Our basic objective was to extend the work of Müller and Trayanov [4] and Gong [22] on pure gluon fields by carrying out studies involving both the quark and the gluon sector simultaneously. For this purpose, we followed the approach suggested by the work of Wong [9] and Elze and Heinz [23]. In this approach one follows the equation of motion of a classical coloured parton in the mean fields generated by the colour current density in the Yang–Mills equations, where the moving coloured partons themselves self-consistently determine the colour current density. Furthermore, in this description, incoherent small wavelength gluons and quarks/antiquarks are treated on the same footing, both being treated as coloured partons. The separation of the gluons into two scales viz. one corresponding to coloured partons and the other corresponding to mean fields is in the nature of an approximation [23]. Compared to full Yang–Mills solutions, it amounts to losing some information and phenomena related to intermediate scales. However, the specific advantage one gets is that one can simultaneously incorporate the contribution of quarks/antiquarks and gluons to the collective thermalization process.

The equations of motion of a classical coloured particle moving in an external colour gauge field were first obtained by Wong [9] from the QCD Lagrangian. It was later pointed out that the equation of motion of the short wavelength gluons can also be described by these equations [23]. The equations of motion of a classical coloured particle moving on a world line $x^\mu(\tau)$ and having momentum $p^\mu = m(\mathrm{d}x^\mu/\mathrm{d}\tau)$ are given by

$$\frac{\mathrm{d}p^\mu}{\mathrm{d}\tau} = gI_a v_\nu F_a^{\mu\nu}, \tag{2}$$

$$\frac{\mathrm{d}I_a}{\mathrm{d}\tau} = -gf^{abc} v_\mu A_b^\mu I_c, \tag{3}$$

where τ is the proper time of the particle. The gauge field $F_a^{\mu\nu}$ and the potential A_a^μ correspond to the mean fields and are evaluated at point $x^\mu(\tau)$. g is the coupling constant and f^{abc} are the structure constants of the relevant gauge group. I_a is an N^2-1 component vector of the colour charge through which a coloured particle couples with the gauge fields. In fact, (2) is very similar to the Lorentz force equation of the electromagnetic theory. Equation (3) describes the dynamics of the colour charge of the particle, which has no analogue in the electromagnetic theory. The meaning of classical colour charge of a particle becomes clear if it is interpreted in terms of the expectation value of the quantum current operator $j_{\mathrm{op}}^{\mu a}(x) = g\bar{\psi}(x)\gamma^\mu\tau^a\psi(x)$. One can carry out the identification

$$\begin{aligned} j_{\text{classical}}^{\mu a}(x) &= g\int \mathrm{d}\tau v^\mu \delta^4(x - x(\tau))I^a(\tau) \\ &\equiv g\langle\bar{\psi}(x)\gamma^\mu\tau^a\psi(x)\rangle. \end{aligned}$$

It should be mentioned that $I_a(\tau)$ is a continuum variable and hence is strictly speaking appropriate in the large N limit. The gauge field tensor $F_a^{\mu\nu}$ is given by

$$F_a^{\mu\nu} = \partial^\mu A_a^\nu - \partial^\nu A_a^\mu + gf_{abc}A_b^\mu A_c^\nu. \tag{4}$$

Dynamics of mean field $F_a^{\mu\nu}$ is described by the Yang–Mills equations

$$\partial_\mu F_a^{\mu\nu} + gf_{abc}A_{\mu b}F_c^{\mu\nu}, = j_a^\nu, \tag{5}$$

where j_a^ν are the components of the current density. j_a^ν can be evaluated from the knowledge of the coloured particle trajectories obtained from (2) and (3). In general, it is very difficult to solve (2)–(5) for a many particle system. In order to gain insight into the collective modes described by such a system, we first study them in the hydrodynamic limit. Equations describing the colour hydrodynamics were first written by Kajantie and Montonen [24], based on (2)–(5). Such equations can be obtained by taking the moments over the many particle distribution (see for example Bhatt [5]).

4. Colour hydrodynamics approach to collective behaviour

The basic idea in this approach is to study collective properties of QGP in a classical but non-perturbative manner. For this purpose we use the colour hydrodynamic equations deduced by Kajantie and Montonen [24] from Wong's equations discussed in the last section.

We would like to mention that the classical approach has been applied to the study of collective properties in quark gluon plasmas by other workers also [25]. In these papers, however, the plasma was considered to be essentially abelian in nature with perturbations taken in a linear form. In the work carried out by our group the non-abelian character of the plasma is retained [6, 7]. The hydrodynamical frame work was preferred over the kinetic or other detailed microscopic models because hydrodynamic models are simpler to deal with and they can describe almost all the collective modes described by the kinetic theory. Further, for simplicity, colour $SU(2)$ group is chosen instead of $SU(3)$.

The colour hydrodynamic (CHD) equations describe the space-time evolution of classical coloured particles in the fluid picture. These equations contain three continuous quantities: number density of the coloured particles n, velocity field $\mathbf{V}$ and the colour charge vector $I_a(a=1,2,3)$ which are functions of space and time. Their dynamics is described by [24] the equations

$$\partial_t n_A + \nabla\cdot(n_A\mathbf{V}_A) = 0, \tag{6}$$

$$\partial_t\mathbf{V}_A + \mathbf{V}_A\cdot\nabla\mathbf{V}_A = (g/m_A)\sqrt{1-V_A^2}I_{Aa}[\mathbf{E}_a + \mathbf{V}_A\times\mathbf{B}_a - \mathbf{V}_A(\mathbf{V}_A\cdot\mathbf{E}_a)], \tag{7}$$

$$\partial_t\mathbf{I}_{Aa} + \mathbf{V}_A\cdot\nabla\mathbf{I}_{Aa} = -g\epsilon_{abc}[A_b^0 - \mathbf{V}_A\cdot\mathbf{A}_b]\mathbf{I}_{Ac}. \tag{8}$$

In (6)–(8) the suffix A and m_A denote species label of coloured particles and mass of species A respectively. For light relativistic particles mass m_A is replaced by (enthalpy density/number density) for the relevant specie fluid. ϵ_{abc} is the completely antisymmetric Levi–Civita tensor and $A_a^\mu(\mu=0,1,2,3)$ are $SU(2)$ gauge potentials which satisfy Yang–Mills equations $\partial_\mu F_a^{\mu\nu} + g\epsilon_{abc}A_{\mu b}F_c^{\mu\nu} = j_a^\nu$. Here, j_a^ν represents matter current which can be expressed in terms of CHD variables as $j_a^0 = g\sum_A n_A I_{Aa}$ and $\mathbf{j}_a = g\sum_A n_A\mathbf{V}_A\mathbf{I}_{Aa}$. Equations (6)–(8) together with Yang–Mills equations form a closed set of Lorentz and gauge covariant equations [6]: which describe the evolution of the QGP self consistently. It should be noted that among the CHD variables n_A and $\mathbf{V_A}$ are gauge invariant quantities while the colour vector $\mathbf{I}_{Aa}$ transforms gauge covariantly [6].

Our model does not explicitly take account of the effects due to pair creation and annihilation processes $q\bar{q}\leftrightarrow g$ and $g\bar{g}\leftrightarrow g$. This would lead to appropriate source and sink terms in the equations of continuity, motion, etc. We now show that for collective phenomena, this is justifiable provided the collision-frequency for momentum transfer due to collective effects is larger than the appropriate pair creation/annihilation collision frequencies. The typical ratio of the plasma frequency ω_p to the collision frequency ν of pair creation/annihilation processes may be estimated as [25] $\omega_\mathrm{p}/\nu = 4\pi/\alpha_s\sqrt{(N_f+2N)/18}\,(\ln(\alpha_s^{-1}))^{-1}$. Here, N_f and N are number of quark flavours and the number of the gluon colours respectively. g is the strong interaction coupling constant and $\alpha_s = g^2/4\pi$. For the plasma with three flavours of quarks and $g\sim 1$, say at temperature $\sim 200\,\mathrm{MeV}$, one finds $\omega_\mathrm{p}/\nu\sim 10$. Thus one may have significant momentum transfer by collective processes compared to the pair creation/annihilation collisional processes and the latter may be neglected. In our model, then, the equilibrium number densities of partons are fixed and are prescribed externally, based on our knowledge of the background temperature, energy density etc.

We now use (6)–(8) to describe thermalization due to non-abelian collective effects in a quark gluon plasma. Let us imagine that because of the violent nature of a heavy ion collision a significant energy is stored in coherent electric fields. For simplcity, we assume that most of this energy is in the form of longitudinal oscillations. We would like to address the question: what happens to the energy as the system evolves? Is there any evidence that this energy may be thermalizing? If so, what are the time-scales involved?

To investigate these questions, we have studied non-abelian longitudinal oscillations of a quark gluon plasma. If the colour electric field is produced in the z-direction, then, we can simplify (6)–(8) and the YM equations by choosing a purely longitudinal mode (i.e., $A_a^x = A_a^y = 0$) and choosing the gauge $A_a^0 = 0$. The potentials A_a^z are functions of z and t variables, but we further assume that they depend only on the variable $\zeta = z + \beta t$. Physically this means that one is studying a class of solutions which are stationary in a frame moving with phase velocity β. Such stationary wave solutions are widely studied in the Coulomb plasma literature and have also been considered for the non-abelian theories [26, 27]. With these assumptions the basic CHD equations reduce to three coupled non-linear second order ordinary differential equations having a dimensionless parameter $\epsilon = g^2 I_0 a_0^2/m\omega_\mathrm{p}$ where, I_0 and a_0 represent the scaling parameters for the colour charge and the gauge field amplitude A_a^z respectively and ω_p is the plasma frequency. We next

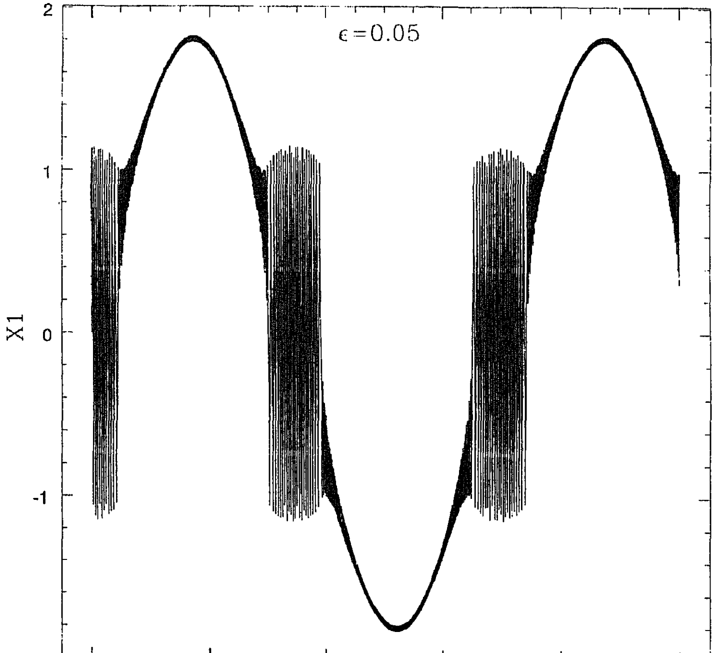


Figure 1. Hydrodynamic variable $X1$ vs. time for non-abelian parameter $\epsilon = 0.05$.

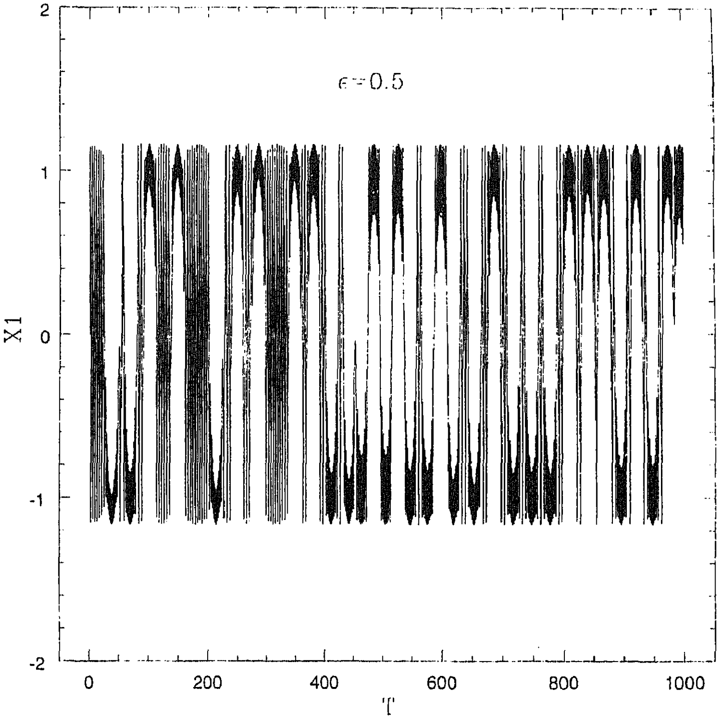


Figure 2. Hydrodynamic variable $X1$ vs. time for non-abelian parameter $\epsilon = 0.5$.

define quantities $X_i (i = 1, 2, 3)$, which are linear combinations of the scaled gauge potential $A_a^z \equiv A_a a_o$. They are $X_1 = A_1 + A_2 + A_3$, $X_2 = \sqrt{3/2}(A_1 - A_2)$ and $X_3 = \sqrt{1/2}(A_1 - 2A_2 + A_3)$. In terms of the X_i's the equations reduce to

$$\ddot{X}_1 = -3X_1 + (\epsilon/\sqrt{3})(X_2\dot{X}_3 - X_3\dot{X}_2)X_1, \tag{9}$$

$$\ddot{X}_2 = (\epsilon/\sqrt{3})(X_3\dot{X}_1 - X_1\dot{X}_3)X_1, \tag{10}$$

$$\ddot{X}_3 = (\epsilon/\sqrt{3})(X_1\dot{X}_2 - X_2\dot{X}_1)X_1. \tag{11}$$

We would like to note that for the case of homogeneous plasma i.e. $k \longrightarrow 0$, we get the same set of equations as that with the stationary frame ansatz (for details of both calculations, see Bhatt *et al* [6]). Our main findings are as follows: (i) for the case $\epsilon = 0$ we get the longitudinal plasma oscillations with constant amplitude and frequency, which are very similar to those in the Coulomb plasma. In this case, in fact, the equation describing the oscillation becomes linear. (ii) For small values of $\epsilon (= 0.05)$ we find a new "earphone" oscillation (figure 1) mode in which the high frequency oscillations are modulated at a lower frequency and the abelian oscillation of the plasma at the plasma frequency is periodically followed by a non-abelian oscillation with higher frequency and reduced amplitude. For higher values of $\epsilon \sim 0.5$ the oscillations are found to exhibit chaotic behaviour (figure 2). We have also studied auto-correlations for such a time series [5]. The analysis shows that the overall decay in time t_c is such that $\omega_p t_c < \omega_p/\nu \sim 10$ as discussed earlier in the section. Thus the decay time scale for oscillations can be faster than the collisional time scale for larger values of the parameter ϵ. We have found from this analysis that non-abelian effects lead to chaos in plasma oscillations. Such chaotic phenomena are already observed in the vacuum solutions of the Yang–Mills theory.

The hydrodynamic appraoch, however, does not give any information about the evolution of parton distribution function. Such information can be obtained from the solutions of (2)–(5). The chaotic behaviour of the plasma oscillations have implication for the thermalization of the energy. To analyze this question in more detail we have carried out a $1+1$ dimension particle simulation of the longitudinal oscillations in the QGP.

5. Particle simulations of a coloured parton plasma

Particle simulations are extensively used to study the nonlinear physics of classical electrodynamic plasmas. These simulations are in the nature of computer experiments in which one explicitly follows the motion of N particles in their own self-consistent fields which are evolved using the Maxwell field equations. Since the simulations are typically done to explore collective nonlinear phenomena, one usually discriminates against short range Coulomb interactions (collisional effects) by using large number of particles in a Debye sphere ($\Lambda_{\mathrm{D}} \gg 1$) and/or by using smoothed out 'fat' particles [28]. Due to limitations imposed by computers, particle simulations are typically restricted in maximum N usable, dimensionality and noise effects/errors due to finite grid size and time step of integration. Within these limitations, particle simulations can go well beyond the hydrodynamic description of last section and can give us a detailed x, t evolution of fields, distribution function of particles in phase space and so on. It is therefore particularly well suited for investigating questions relating to collective thermalization in coloured parton plasmas.

We have written a relativistic particle simulation code for colour partons [8]. In the code we follow the motion of a collection of classical coloured partons interacting via non-abelian gauge fields. The classical partons (which represent quarks and short scale gluons as discussed in § 3) constituting the plasma are described using three variables: the particle position **x**, velocity **v** and the colour charge I_{Aa}. The dynamics of the particles is governed by the relativistic Lorentz force law due to Wong [9] viz. (2). The dynamics of the fields $F_a^{\mu\nu}$ is given by the Yang–Mills equation, (5). The coloured partons interact by generating a charge/current density j_a^ν which acts as a source for the Yang–Mills fields. These fields exert force on the partons and also rotate their colour vectors. The variation of colour charge of a particle is governed by the colour dynamic equation, (3). It is this attribute of a coloured parton plasma which distinguishes it from conventional electrodynamic plasmas and makes our simulation, first one of its kind. The equilibrium densities of partons are treated as given constants (since we are not explicitly incorporating effects due to particle creation and annihilation). Finally, the equilibrium parton densities are distributed in such a manner that the plasma has overall colour neutrality, as is expected in a parton plasma created by heavy ion collisions.

Our objective in these simulations is to investigate collective aspects of thermalization in non-abelian plasmas. We thus start with a cold ($T \to 0$) coloured parton plasma with large Λ_D and at $t = 0$ put a lot of energy in coherent longitudinal fields (such as the ones which might be created because of heavy ion collisions, as discussed in § 4). Restricting ourselves (for simplicity) to longitudinal field interactions only we ask, how does this energy equilibrate and what is the time scale of equilibration? For this purpose we study the time evolution of fields (where we find generation of chaos and energy spreading in Fourier space) and parton distribution functions (where we find heating and evolution to Maxwell/Maxwell–Jüttner distributions). Finally, we ask whether in the thermalized state, there is an equipartition of energy between the mean field modes and the kinetic energy per parton (and we find it is indeed so!).

As discussed above, we restrict the simulations to dynamics of longitudinal colour fields only and work in $1 + 1$ dimensions. We use the gauge condition $A_a^1 = 0$ so that our basic equations take the form

$$\frac{\mathrm{d}\gamma x_A}{\mathrm{d}t} = v_A, \tag{12}$$

$$\frac{\mathrm{d}\gamma v_A}{\mathrm{d}t} = I_{Aa} E_a, \tag{13}$$

$$\frac{\partial E_a}{\partial x} = n_A I_{Aa}, \tag{14}$$

$$\frac{\partial A_a^0}{\partial x} = -E_a, \tag{15}$$

$$\frac{\mathrm{d}I_{Aa}}{\mathrm{d}t} = -\alpha \epsilon_{abc} A_{0b} I_{Ac}, \tag{16}$$

where $\gamma_A \equiv (1 - v_A^2/c^2)^{-1/2}$, A refers to parton species, repeated indices are to be summed over and we have made the equations dimensionless with the following scalings: $x_A \rightarrow x_A/k_L^{-1}, v_A \rightarrow v_A/\omega_p k^{-1}$, $t \rightarrow \omega_p t, I_{Aa} \rightarrow I_{Aa}/I_0$, $n_A \rightarrow n_A/n_0, E_a \rightarrow E_a/(g I_0 n_0 k^{-1} \hbar^{-1})$, $A_a^0 \rightarrow A_a^0/(g I_0 n_0 k^{-2} \hbar^{-1})$; here $\omega_p^2 = g^2 I_0^2 n_0 / m\hbar^2$ is the plasma frequency, $\alpha = m\omega_p / I_0 k^2$ is the non-abelian parameter and $k_L = 2\pi/L$ is the wave-vector of the lowest mode.

We use the particle-in-cell (PIC) method of coloured parton simulations. An infinite 1-D system is modelled using periodic boundary conditions. The whole system is divided into NG cells using $NG + 1$ grid points and all the field equations like electric field, potential, charge density etc., are calculated at the grid points. The particles are initially positioned at the centre of the cells and each particle is assigned a zero initial velocity (cold parton plasma with $T = 0$) and 3-components of colour charge. Particles are initially given a perturbation in colour space to set up the colour charge density and the associated longitudinal colour fields. Once the initial conditions are given, the charge density on the grid points are calculated using a weighting scheme, which implies some form of interpolation among the grid points nearest to a particle. We use a quadratic spline weighting scheme in which one distributes the charge of a particle to three nearest grid points using the assignment functions

$$W(X_j - x) = \frac{1}{\Delta x}\left[\frac{3}{4} - \left(\frac{x - X_j}{\Delta x}\right)^2\right]$$

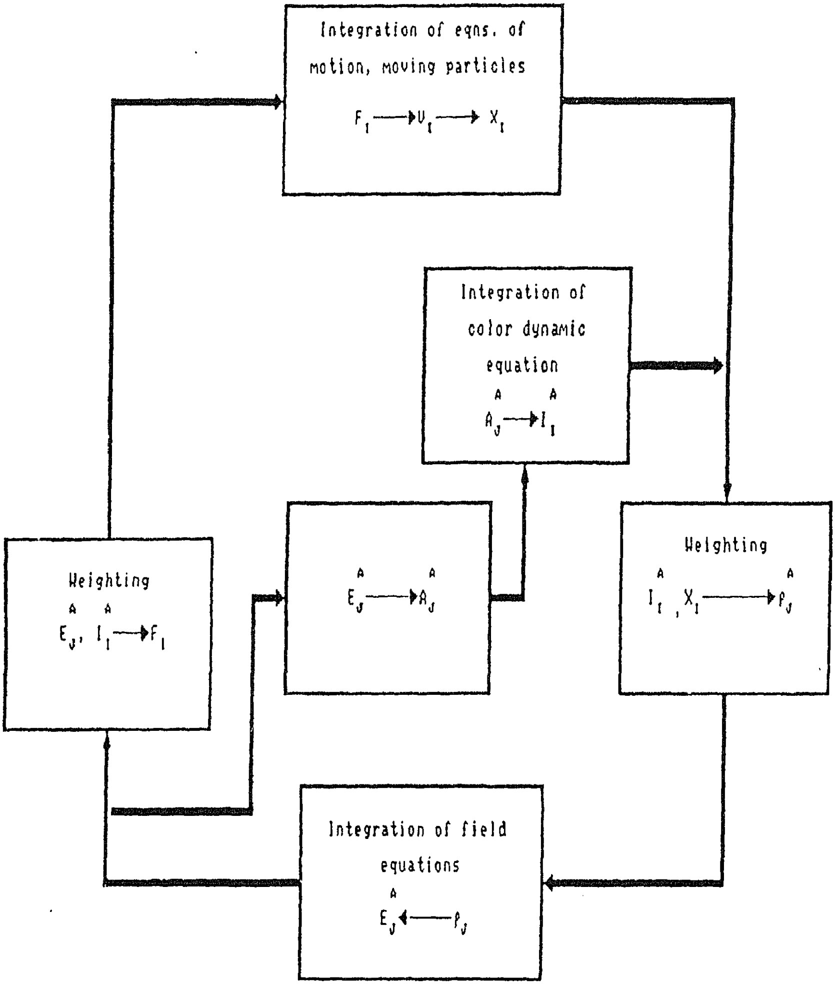

Integration of eqns. of motion, moving particles
$F_I \rightarrow v_I \rightarrow X_I$
Integration of color dynamic equation
$\hat{A}_J \rightarrow \hat{I}_I$
Weighting
$\hat{E}_J, \hat{I}_I \rightarrow F_I$
$\hat{E}_J \rightarrow \hat{A}_J$
Weighting
$\hat{I}_I, X_I \rightarrow \hat{\rho}_J$
Integration of field equations
$\hat{E}_J \leftarrow \rho_J$

Figure 3. Computational cycle; $I \rightarrow$ particle index, $J \rightarrow$ grid points, $A \rightarrow$ colour index.

and

$$W(X_{j\pm1} - x) = \frac{1}{2\Delta x}\left[\frac{1}{2} \pm \frac{x - X_j}{\Delta x}\right]^2,$$

where X_j's are the grid coordinates and x is the coordinate of a particle. This process of distributing the charge of a particle to nearest three grid points is repeated for all the particles and so the total charge density at a particular grid point is the sum of the contribution from all particles contained within a width $3\Delta x$ with that grid point at the centre. This description should also give one an idea of how the 'fat particle' concept, suppressing the short range 'collisional' Coulomb interactions is introduced in the simulations.

The next step in simulation is to solve the Gauss law on the grid. We use the simplest central difference scheme accurate up to order $(\Delta x)^2$ for this purpose. To make the average field zero, we calculate $\langle E \rangle$ and subtract it from each E_j. The potentials required for solving the colour dynamic equation are also calculated in this step by using a similar central difference scheme. After the electric field calculation, the forces acting on the particles are calculated. To calculate the force, one needs to calculate the electric fields at the particle positions. For this purpose, we use a similar interpolation scheme as before. This is important because it can be shown [29] that for identical charge assignment and force interpolation schemes and correctly space-centred difference approximations for derivatives, the self force on a particle arising due to various particle to grid and grid to particle interpolations is zero and the total momentum of the system is identically conserved. In the next step, forces calculated on the particles are used to move the particles to a new position by solving the equations of motion. We use the standard leap-frog scheme [30] for this purpose. The last step in the computational cycle is the calculation of colour vector for each particle by solving the colour dynamic equation. We have used a predictor-corrector scheme [see e.g. 31] for this purpose. All the steps used in a computation cycle are summarised in the box diagram shown in figure 3.

We now report the main results of our simulations. As discussed above, we initiate the simulation with the imposition (at $t = 0$) of a coherent colour charge perturbation $I_a = I_{a0}(1 + \delta \cos(kx + \theta_a))$ and study the time evolution of the system for different values of the non-abelian parameter α. We choose 1024 grid points in a system of length 2π and put 10^4 particles on it. The parameter $n\lambda_{\mathrm{D}}$ for the 1-D system $\simeq 13$ (which roughly corresponds to a 3-d parameter $\Lambda_{\mathrm{D}} \sim 2 \times 10^3$). We use relatively small values of δ so that the electric field fluctuations induced by colour charge density are small; this

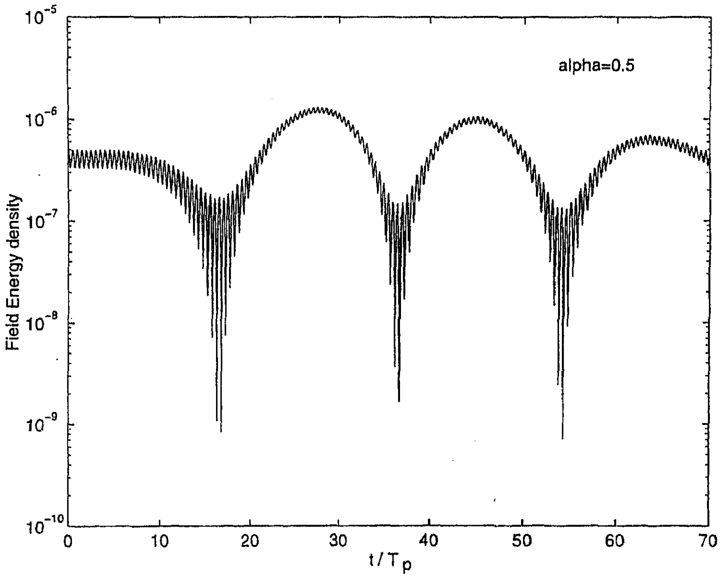


Figure 4. Field energy density vs time for non-abelian parameter $\alpha = 0.5$.

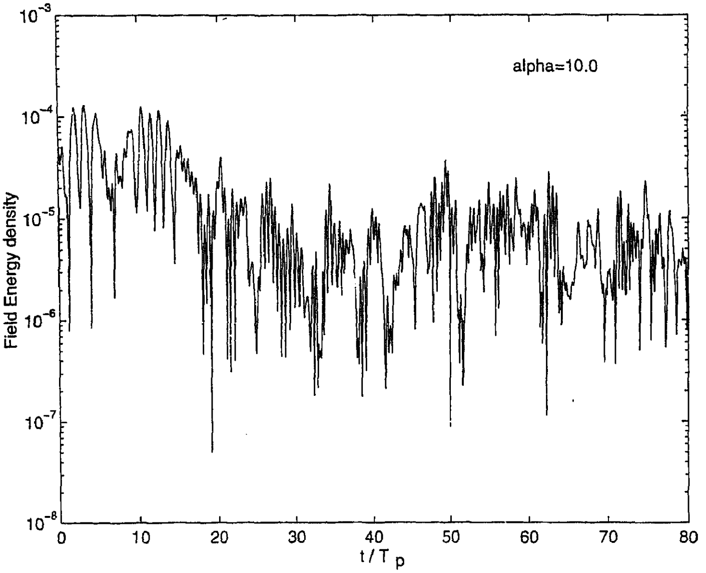


Figure 5. Field energy density vs time for non-abelian parameter $\alpha = 10.0$.

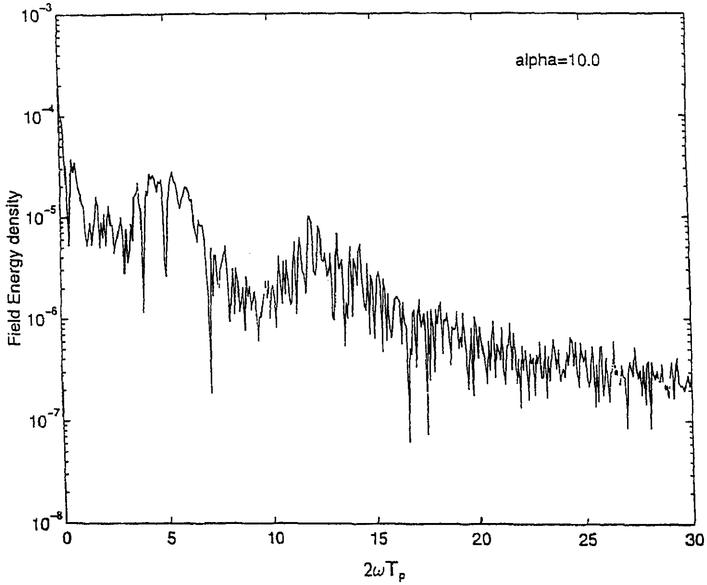


Figure 6. Field energy density vs frequency (power spectrum) for non-abelian parameter $\alpha = 10.0$.

is done so that the effects due to conventional plasma nonlinearity (coming, for example from the convective terms in hydrodynamic equations) are negligible. We then control the strength of non-abelian interactions by changing the parameter α, which essentially determines the rate of precession of the colour charge vector. It should be emphasized that in an actual parton plasma, the product $\alpha\delta$ determines the rate of precession of the colour vector so that this may be enhanced by changing the amplitude of initial perturbation also and furthermore the non-abelian and plasma nonlinearities both operate together, contributing to the overall collective thermalization process.

For $\alpha = 0$, we find that the system exibits usual plasma oscillations at the plasma frequency. This is expected since the non-abelian physics is absent and conventional plasma nonlinearities are negligible(as explained above). Figure 4 shows the behaviour of collective oscillations for $\alpha = 0.5$; we have chosen field energy density at a particular point as a representative variable and show its time evolution. The system continues to be coherent and exhibits characteristic 'ear phone' oscillations in which a slower frequency modulates the usual plasma oscillations [9]. It is interesting that the full particle simulation reproduces a behaviour observed in simple hydrodynamic models (figure 1).

Figure 5 shows the characteristic time evolution for $\alpha = 10.0$. Now the field energy density shows irregular chaotic oscillations (similar to figure 2 in § 4). To examine the nature of this chaos we plot the power spectrum in figure 6. We see a broadband structure which is fairly representative of chaotic systems but this by itself does not guarantee that the system exhibits sensitivity to initial conditions [32]. To examine this question we define a gauge invariant metric as

$$D(t) = \sqrt{\sum_{i=0}^{NG-1} [E_a^2 - E_a'^2]^2} \tag{17}$$

This is nothing but the distance between two points in a NG-dimensional space, where each point represents the field energy configuration of the whole system. A similar type of definition has been used by Müller *et al* [4] and Gong *et al* [22] to exhibit chaotic dynamics in a pure gluon system. In figure 7 we show the variation of $\ln(D(t))$ versus time. The linear rise of $\ln(D(t))$ as a function of t shows that the 'distance' between two neighbouring configurations increases exponentially in time. The saturation, as also observed in the earlier work of Müller *et al* (1992), is because of the compact space used for description of fields (viz. finite number of discrete modes). The slope of the linear part gives the largest Lyapunov exponent (h) which is related to the rate of entropy rise [4, 31, 34]. The inverse of the Lyapunov exponent gives an estimate of the entropy production time scale or the thermalization time. In our case the estimated thermalization time turns out to be of order 1.4 plasma periods. We shall discuss the notion of thermalization in somewhat more detail below.

We first look at the distribution of energy in the various Fourier modes after the elapse of several Lyapunov exponent times. Figure 8 shows the Fourier mode energy density vs. k/k_L. We see that energy is distributed in a large number of modes although initially (at $t = 0$) we had pumped energy only into the lowest mode. At the same time we look at the phase space of partons. Compared to the initial cold $(T = 0)$ parton distribution, we observe a considerably broadened phase space (figure 9). The space average distribution of particle momenta is shown in figure 10. It shows that the distribution approaches a Maxwellian form with a temperature of order $T/mc^2 \sim 10^{-4}$. Interestingly, we also find that mean energy per mode is of order $T/2$ as would be expected from the law of equipartition of energy in a thermalized distribution.

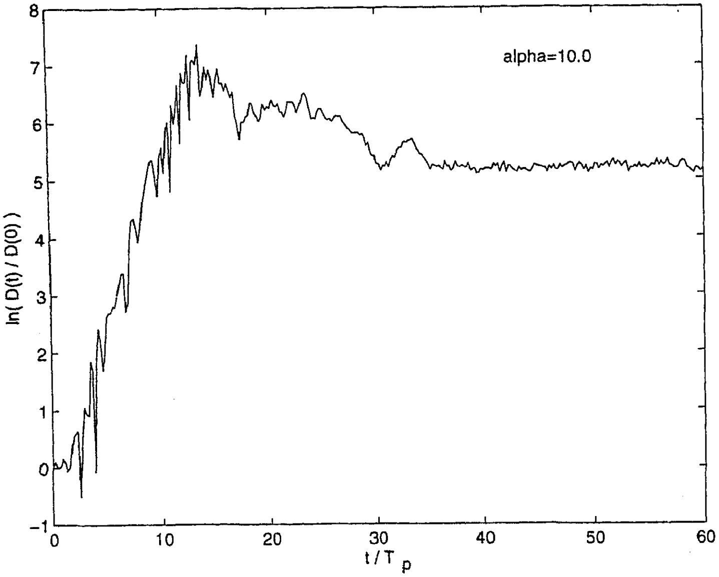


Figure 7. $\ln(D(t))$ vs time for non-abelian parameter $\alpha = 10.0$.

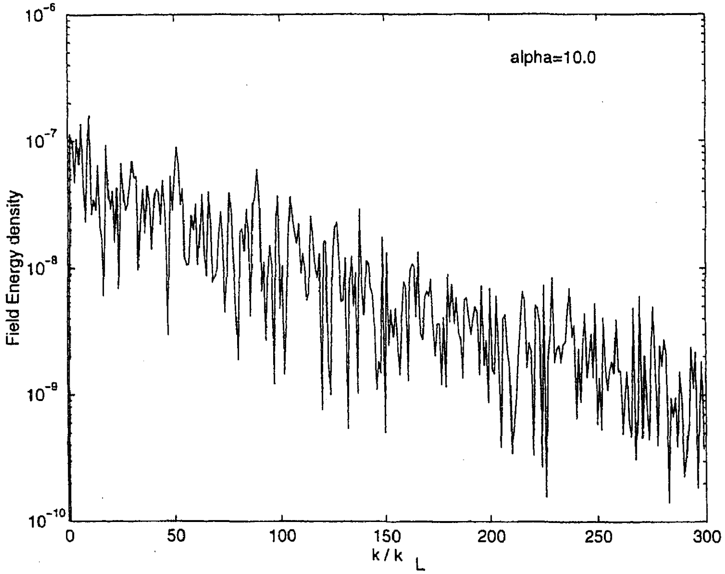


Figure 8. Field energy density vs mode number for non-abelian parameter $\alpha = 10.0$ at normalized time $t/T_p = 80$.

We now comment on various aspects of the observed thermalization process. The flow of energy from low k to high k values is due to non-abelian interactions and is expected in a classical system tending towards energy equipartition. The Rayleigh–Jeans distribution

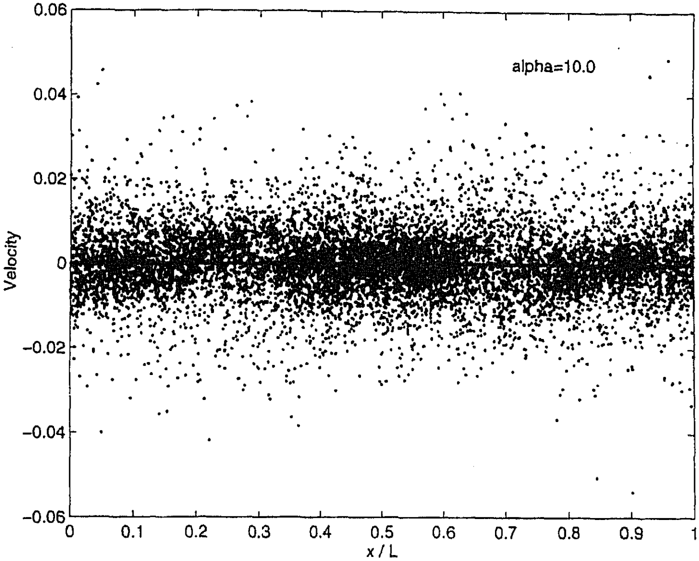


Figure 9. Phase space of partons for non-abelian parameter $\alpha = 10.0$ at normalized time $t/T_p = 80$.

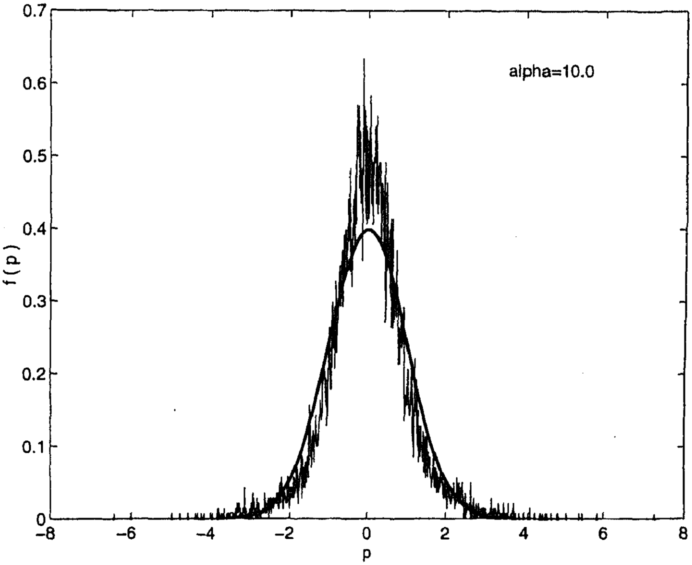


Figure 10. Normalized distribution function for non-abelian parameter $\alpha = 10.0$ at normalized time $t/T_p = 80$.

in a 1-D system would correspond to a flat energy distribution in k space. We have not observed the fully thermalized (i.e. flat!) mean field distribution (figure 8). This may be realization chosen by us. The average 'mean energy' per mode does however appear to be close to $(1/2)T$ as expected in a thermalized distribution.

The parton distribution function however appears to be fully thermalized, having acquired the Maxwell–Boltzmann form shown in figure 10. The mechanism by which cold plasma particles get heated up is related to 'phase mixing' type damping of the longitudinal fields. One may argue that when the initial energy is distributed into a large number of randomly phased Fourier modes, resonant particle effects associated with the high k modes $(\omega/k \sim v_{\mathrm{th}})$ can lead to randomization of particle velocities. The distribution has acquired the Maxwell–Boltzmann rather than the relativistic Maxwell-Jüttner form because we have chosen rather small amplitudes of the longitudinal colour fields. This has restricted the final distribution to non-relativistic temperatures. This was done because we wanted to avoid effects due to plasma nonlinearity and wanted to isolate the physics of non-abelian effects. As mentioned earlier, in a true situation, with larger initial fluctuations, both nonlinearities would contribute to phase mixing and thermalization and lead to Maxwell–Jüttner distribution. It should be emphasised that the thermalized parton distribution does not distinguish between quarks and short scale gluons and both will acquire Maxwell–Jüttner type distributions relativistically. This is because we are treating the short scale gluons as classical distinguishable particles and hence their statistics will only generate the most probable Maxwell–Jüttner form (rather than the Rayleigh–Jeans form which is the $\hbar \rightarrow 0$ limit of the Planck distribution). The essential point is that all distributions have nearly thermalized and taken the most probable form, within the constraints of the model, and everything happens because of collective effects within a few Lyapunov exponent times.

We now comment on the thermalization time observed by us. From the above discussion it appears that although the Lyapunov exponent time is only about 1.4 plasma periods, it takes several plasma periods (12 to 13 in our specific example) to see complete thermalization. We suspect that this may be because of the special initial condition chosen (a coherent longitudinal field plus $T = 0$ parton distributions) and because of the rather restricted nature of interaction in a 1-D plasma with a finite number of discretized modes. If the system exhibits a chaotic attractor (as it does in our case) the time taken for complete thermalization can depend on the 'distance' between the initial state and the final thermalized state. We expect that for general initial realizations and for a system which has a greater number of degrees of freedom either because of three dimensionality or because of greater number of particles/number of modes, the system should approach a thermalized state very soon after arriving on the attractor. We thus take a Lyapunov exponent time as a more representative 'minimum time for thermalization' needed by collective effects studied by us.

In the final section, we present an application of the above results to the thermalization problem in relativistic heavy ion collision.

6. Discussion

The main emphasis in this paper has been on collective aspects of thermalization in a quark gluon plasma. After presenting a summary of the earlier work of Müller *et al* [4] and Gong [22] on thermalization in pure gluon fields, we presented a detailed survey of our work on collective thermalization in classical parton models [6, 8]. The basic conclusion is that when collective effects dominate (typically, when $\Lambda_D \gg 1$), the thermalization time is of order of a plasma period. This time for a typical relativistic heavy ion collision is of order 1 fm/c.

A few words about the limits of validity of our model and its applicability to the RHIC problem are in order here. The best estimates of Λ_D for typical RHIC parameters give $\Lambda_D \simeq 1$ [2], which argues that short range collisional effects are important (giving a thermalization time $\sim$ 1 fm/c) and collective effects may be ignorable. However, there is uncertainity in these estimates and our calculations show that even when $\Lambda_D \gg 1$ and collective effects dominate, thermalization occurs on a time scale $\sim$ 1 fm/c. This is reassuring. Furthermore, when $\Lambda_D \approx 1$, the plasma can go into a strongly coupled state [33] . None of the theoretical calculations (either collective or collisional) presently take this into account. It would be of interest to investigate thermalization in such strongly correlated plasmas. Our simulations are presently restricted to $SU(2)$ group symmetries, longitudinal field interactions and non-relativistic amplitudes only. However, we suspect that the basic result that thermalization time is of the order of a plasma period is unlikely to change since that is the characteristic time associated with collective effects. Thus we can use the results of our simulation for the more general cases also, taking care to define the plasma period in an appropriate manner. The above estimate of 1 fm/c for the plasma period has been obtained in this manner only. We may wonder whether our 'periodic' simulations are truly representative of the finite heavy ion collision experiment. In a way, we have restricted the maximum size of the system by defining a minimum k (or the largest wavelength) in the system. However, our results are relatively independent of this parameter as long as, this largest wavelength is many Debye lengths long. Thus we expect that in the range of dominance of collective effects, our simulation results should be representative.

Finally, a word of caution should be made regarding the application of thermalization results based on classical chaos theory to a truly quantal problem. We hope to address these and related questions in our future work.

References

[1] H Satz, in *Quark matter'95, Nucl. Phys.* **A590**, 63 (1995)
[2] E Shuryak, *Phys. Rev. Lett.* **68**, 3270 (1992)
[3] K Geiger and B Müller, *Nucl. Phys.* **B369**, 600 (1992)
[4] B Müller and A Trayanov, *Phys. Rev. Lett.* **68**, 3387 (1992)
[5] J R Bhatt, *Nonlinear waves in a quark-gluon plasma*, Ph.D Thesis (M S University Baroda, 1992)
[6] J R Bhatt, P K Kaw and J C Parikh, *Phys. Rev.* **D39**, 646 (1989)
[7a] J R Bhatt, P K Kaw and J C Parikh, *Pramana – J. Phys.* **43**, 307 (1994)
[7b] J R Bhatt, P K Kaw and J C Parikh, *Pramana – J. Phys.* **43**, 467 (1994)
[8] S Sengupta, J R Bhatt, P K Kaw and J C Parikh, in preparation (1996)
[9] S K Wong, *Nuovo Cimento* **A65**, 689 (1970)
[10] K Sailer, W Greiner and B Müller, in *Quark-gluon plasma* edited by R Hwa, (World-Scientific, Singapore, 1990) p. 299
[11] M Gyulassy, in *Quark-gluon plasma*, edited by R Hwa, (World-scientific, Singapore, 1990) p. 223

[12] A Casher, H Neuberger and S Nussinov, *Phys. Rev.* **20**, 179 (1979)
[13] A Bialas and W Czycz, *Phys. Rev.* **D30**, 2371 (1984)
[14] J D Bjorken, in *Current-induced reactions* edited by J G Körner, G Kramer and D Schildknecht (Springer-Verlag, Berlin, 1976)
[15] J Schwinger, *Phys. Rev.* **82**, 664 (1951)
[16] A K Ganguly, P K Kaw and J C Parikh, *Phys. Rev.* **D48**, R2983 (1993)
[17] B Banerjee, R S Bhalerao and V Ravishanker, *Phys. Lett.* **B224**, 16 (1989)
[18] M Asakawa and T Matsui, *Phys. Rev.* **43**, 2871 (1991)
[19] K J Eskola and X N Wang, *Phys. Rev.* **D49**, 1284 (1994)
[20] L Xiong and E Shuryak, *Phys. Rev.* **C49**, 2207 (1994)
[21] S C Matinyan, G K Savvidy and N G Ter-Arutyunyan-Savvidy, *Sov. Phys. JETP* **53**, 421 (1981)
[22] C Gong, *Phys. Lett.* **B298**, 257 (1993)
[23] H T Elze and U Heinz, in *Quark-gluon plasma* edited by R Hwa, (World-Scientific, Singapore, 1990) p. 117
[24] K Kajantie and C Montonen, *Phys. Scr.* **22**, 555 (1981)
[25] S Mrówczyńsky, in *Quark-gluon plasma* edited by R Hwa, (World-Scientific, Singapore, 1990) p. 185
[26] A Akhiezer *et al*, in *Plasma electrodynamics, non-linear theory and fluctuations*, part II (Pergamon Press, London, 1975)
[27] J Ishida and A Hosoya, *Lett. Nuovo Cimento* **13**, 237 (1975)
[28] J M Dawson, *Rev. Mod. Phys.* **55**, 2 (1983)
[29] R W Hockney and J W Eastwood, in *Computer simulation using particles* (McGraw Hill, 1981)
[30] C K Birdsall and A B Langdon, in *Plasma physics via computer simulation* (McGraw Hill, 1985)
[31] T Tajima, in *Computational plasma physics: with applications to fusion and astrophysics* (Addison – Wesley Publishing Company, 1989)
[32] G L Baker and J P Gollub, in *Chaotic dynamics: An introduction* (Cambridge University Press, 1990)
[33] S Ichmaru, in *Strongly coupled plasma physics* (North-Holland Delta Series, 1990)
[34] B Müller, in *The proceedings of the pre-equilibrium parton dynamics* edited by X N Wang (Berkeley, University of California, 1993)

Nonlinear propagation of relativistically intense electromagnetic waves in a collisionless plasma

P K KAW and A SEN
Institute for Plasma Research, Bhat, Gandhinagar 382 424, India

Abstract. We discuss the nonlinear propagation of relativistically intense electromagnetic waves into collisionless plasmas with special emphasis on one dimensional plane wave solutions of the propagating, standing and modulated types. These solutions exhibit a rich variety of phenomena associated with relativistic electron mass variation and coupling between transverse electromagnetic and longitudinal fields. They have important applications to problems of laser propagation, self-focusing in overdense plasmas, particle and photon acceleration and to electromagnetic radiation around pulsars.

Keywords. Nonlinear waves; relativistically intense; plasma waves.

PACS Nos 52.35; 52.60

1. Introduction

The propagation of intense electromagnetic waves in a collisionless plasma is a problem of considerable interest because of its applications to laser fusion, modern plasma based particle acceleration methods, laboratory experiments using T^3 (table top terawatt) lasers, low frequency electromagnetic radiation around pulsars etc. In this paper we shall present a general review of work on this topic with special emphasis on the work done by our group.

In a weak electromagnetic wave propagating in a plasma, electrons jitter under the influence of the oscillating electric field vector giving a shielding current which is responsible for the well known dielectric constant expression $\epsilon = 1 - \omega_{\mathrm{p}}^2/\omega^2$ where $\omega_{\mathrm{p}} \equiv (4\pi n_0 e^2/m)^{1/2}$ is the electron plasma frequency (n_0 is the electron number density, m is the electron mass and e is the electron charge). Since the electromagnetic waves are transverse in nature, the electron motion is normal to the propagation vector; this prevents the convective fluid nonlinearities, entering through the $(\mathbf{v}\cdot\nabla)$ terms from becoming effective for the electromagnetic wave. However, as the directed jitter electron velocity $\mathbf{v} \sim e\mathbf{E}/m\omega$ becomes comparable to the velocity of light c, nonlinear effects enter through relativistic mass variation of the electrons and the $\mathbf{v} \times \mathbf{B}$ forces where $\mathbf{B}$ is the magnetic field associated with the electromagnetic wave. The $\mathbf{v} \times \mathbf{B}$ forces drive the electrons along the propagation direction and give the possibility of coupling to longitudinal plasma oscillations. A relativistically intense electromagnetic wave ($v/c \sim 1$) is therefore always nonlinear and in general a mixed wave, with significant space charge fluctuations associated with it. It is only for highly specialized polarizations, such as circular, where the $\mathbf{v} \times \mathbf{B}$ forces exactly cancel that we have a pure nonlinear electromagnetic wave. The possibility of exciting space charge fields propagating nearly at the velocity of light (viz. the group velocity of an electromagnetic pulse) is the key to modern plasma based particle acceleration schemes, which are currently undergoing intense experimental investigation. We may also physically anticipate other interesting consequences of working with relativistically intense electromagnetic waves. It is obvious that as the electron jitter velocity approaches c, the effective mass of the electrons will increase. This would encourage penetration of intense electromagnetic waves into normally overdense plasmas, i.e., $\omega_{\mathrm{p}} > \omega > \omega_{\mathrm{peff}}$. Similarly, if we send a finite plane wave front with an intensity variation on it, the intense regions will propagate slower than the weakly intense regions, because $c/(1-\omega_{\mathrm{peff}}^2/\omega^2)^{1/2} < c/(1-\omega_{\mathrm{p}}^2/\omega^2)^{1/2})$; this causes the plane wave front to acquire curvature and leads to self-focusing and filamentation of the intense electromagnetic wave. We will consider examples of all the above physical effects in our subsequent discussions.

Our paper is organized as follows. We first discuss the basic equations of the plasma model and its range of validity. We next investigate some special solutions of the full nonlinear set of equations. These solutions assume the existence of nonlinear stationary plane waves and reduce the problem to a set of coupled ordinary differential equations. We then go on to discuss some exact solutions for circularly polarized waves and numerical/approximate analytical solutions for linearly polarized waves. Following Akhiezer and Polovin [1] we next show how the linearly polarized waves may be reduced to the problem of a Hamiltonian with two degrees of freedom and describe a few analytical results obtained by them for high phase velocity waves and waves close to c. We then present some of our own numerical results of the problem, which for the first time showed the rich variety of possible solutions [2]. The appropriate Hamiltonian involves square roots because of relativistic effects (unlike most commonly studied nonlinear Hamiltonians which tend to involve polynomials). The results of these investigations are therefore of general interest to nonlinear dynamics. Since the numerical results indicated an almost integrable system there was widespread interest for a while to obtain an analytical confirmation of integrability. It came as somewhat of a surprise when it was actually proved to be nonintegrable [3]. In § 4, we discuss the extension of these results to standing nonlinear waves. The next section is devoted to modulated nonlinear waves in which solutions are not stationary in a moving frame. This allows us to consider the possible excitation of one dimensional solitons in the coupled electromagneticwave-plasma wave problem. Many features of the numerical solutions of this soliton problem can be reproduced analytically using a WKB method. In the last section we summarise the applications of the above results to problems of laser propagation and self-focusing in overdense plasmas, the problems of particle and photon acceleration and to electromagneticfields around rotating pulsars. We also discuss questions related to stability of these solutions, the importance of multidimensional effects and the conditions under which movement of ions can significantly alter the conclusions.

2. Plasma model and basic equations

Consider a finite amplitude electromagnetic wave propagating along the z axis. If the wave amplitude is sufficiently large so that $eE/m\omega c \gg 1$, the electrons are accelerated to the speed of light in a fraction of the wave period. Two nonlinear effects immediately become important, namely, the relativistic variation of the electron mass and the excitation of longitudinal space charge fields by strong $\mathbf{v} \times \mathbf{B}$ forces driving electrons along the direction of propagation. On the time scale of the electron motion the ions are virtually immobile due to their larger mass and hence their dynamics can be ignored. Of course, at larger electric field strengths, this assumption can break down particularly when the ions also acquire large induced directed velocities. Thus one is restricting oneself to wave amplitudes where the electron directed velocity is large but the ion directed velocity is still small ($eE/M\omega \sim 0.1c$ where M is the ion mass). Likewise the electrons can be treated as a cold fluid since their induced velocity is close to the speed of light and therefore the electron thermal velocity can be neglected in comparison. The basic nonlinear propagation equations for the coupled electromagnetic and longitudinal waves can then be written down in terms of the following set of relativistic cold fluid equations.

$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t}, \tag{1}$$

$$\nabla \cdot \mathbf{E} = -4\pi e(n - n_0), \tag{2}$$

$$\nabla \times \mathbf{B} = -\frac{4\pi e}{c} n\mathbf{v} + \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t}, \tag{3}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{4}$$

$$\frac{\partial n}{\partial t} + \nabla \cdot (n\mathbf{v}) = 0, \tag{5}$$

$$\frac{\partial \mathbf{p}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{p} = -e\left(\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c}\right), \tag{6}$$

where $\mathbf{E}$, $\mathbf{B}$ are the electric and magnetic field components of the wave, $\mathbf{v}$ is the fluid velocity, n is the density and $\mathbf{p} = m\mathbf{v}\gamma$ is the relativistic momentum (with $\gamma = 1/\sqrt{1 - v^2/c^2}$). Equations (1–4) are the full set of Maxwell's equations, (5) is the continuity equation and (6) is the relativistic fluid momentum equation. These model equations were first written down by Akhiezer and Polovin [1] and have formed the basis of most fundamental studies related to the propagation of intense electromagnetic waves in a plasma. While a fully time dependent solution of these coupled nonlinear partial differential set of equations is difficult to obtain analytically (or even handle numerically), various special solutions have been obtained from time to time. A large class of them pertain to so called one dimensional traveling wave solutions, which we discuss next.

3. Traveling wave solutions

We restrict ourselves to one dimension (z the direction of propagation) and look for solutions which are functions of the independent variable $\zeta = z - ut$. Such waves, if they exist, would be nonlinear waves that are stationary in a frame moving with the velocity u. We adopt the following normalisations,

$$\beta = \frac{u}{c}; \quad \rho = \frac{\mathbf{p}}{mc}; \quad \zeta = \frac{\omega_p}{c}(z - ut).$$

Then combining (1–6) and eliminating the electric and magnetic fields, we obtain the following set of coupled differential equations:

$$\frac{\mathrm{d}^2\rho_x}{\mathrm{d}\zeta^2}+\frac{1}{(\beta^2-1)}\frac{\beta\rho_x}{\beta(1+\rho^2)^{1/2}-\rho_z}=0, \tag{7}$$

$$\frac{\mathrm{d}^2\rho_y}{\mathrm{d}\zeta^2}+\frac{1}{(\beta^2-1)}\frac{\beta\rho_y}{\beta(1+\rho^2)^{1/2}-\rho_z}=0, \tag{8}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\zeta^2}[\beta\rho_z-(1+\rho^2)^{1/2}]+\frac{\rho_z}{\beta(1+\rho^2)^{1/2}-\rho_z}=0 \tag{9}$$

The coupled equations above are still difficult to solve analytically but some special exact solutions exist.

3.1 *Pure transverse waves*

If we set $\rho_z=0$, then for bounded solutions (9) gives $\rho^2=$ constant. Equations (7, 8) then admit the solution

$$\rho_x=\rho\cos\omega\left(t-\frac{z}{u}\right), \tag{10}$$

$$\rho_y=\rho\sin\omega\left(t-\frac{z}{u}\right), \tag{11}$$

where $\omega=\omega_{\mathrm{p}}\beta(\beta^2-1)^{-1/2}(1+\rho^2)^{-1/4}$. The wave phase velocity u can be expressed in terms of β as

$$u=\beta c=c\epsilon^{-1/2},$$

where ϵ, the dielectric constant of the plasma, is given by

$$\epsilon=1-\frac{\omega_{\mathrm{p}}^2}{\omega^2}(1+\rho^2)^{-1/2}. \tag{12}$$

Substituting for ρ in the Maxwell's equations, it can be shown that

$$E_x=\left(\frac{mc\omega\rho}{e}\right)\sin\left(\frac{\omega}{u}\zeta\right), \tag{13}$$

$$E_y=\left(\frac{mc\omega\rho}{e}\right)\cos\left(\frac{\omega}{u}\zeta\right) \tag{14}$$

so that

$$\rho^2=(e^2E_0^2/m^2c^2\omega^2),$$

where $E_0^2=E_x^2+E_y^2$. Combining these results one can now see that

$$\epsilon=1-\frac{\omega_{\mathrm{p}}^2}{\omega^2}\left(1+\frac{e^2E_0^2}{m^2c^2\omega^2}\right)^{-1/2}. \tag{15}$$

Equations (13) and (14) indicate that only circularly polarized modes may propagate as pure transverse waves in the plasma. Equation (15) demonstrates that such waves can propagate into overdense plasmas provided their frequencies lie in the range

$$\omega_{\mathrm{p}}\left(1+\frac{e^2E_0^2}{m^2c^2\omega^2}\right)^{-1/2} < \omega < \omega_{\mathrm{p}}.$$

We will return to the physical implications of this effect in the final section.

3.2 *Pure longitudinal waves*

For $\rho_x = \rho_y = 0$, (7, 8) are identically satisfied and $\rho_z = \rho$ in (9). This can be integrated exactly to yield a periodic solution which is expressible in terms of elliptic functions. The frequency of the wave is given by

$$\omega = \omega_{\mathrm{p}}\frac{\pi}{\sqrt{2}}\left[\int_0^{\alpha_m}\frac{\mathrm{d}\alpha}{(1-\alpha^2)^{5/4}}\left(\frac{(1-\alpha^2)^{1/2}}{(1-\alpha_m^2)^{1/2}}-1\right)^{-1/2}\right]^{-1}, \tag{16}$$

where α_m is related to the electric field amplitude through Maxwell equations and is given by

$$eE = \sqrt{2}m\omega_{\mathrm{p}}c\left(\frac{1}{(1-\alpha_m^2)^{1/2}}-\frac{1}{(1-\alpha^2)^{1/2})}\right)^{-1/2}. \tag{17}$$

In the limit $\alpha_m \ll 1$, these expressions can be simplified to

$$\omega = \omega_{\mathrm{p}}\left(1-\tfrac{3}{16}\alpha_m^2\right)$$

with $\alpha_m = (eE_0/m\omega_{\mathrm{p}}c)$ where E_0 is the maximum value of E. For $\alpha_m \to 1$, one gets

$$\omega \approx 2^{-3/2}\pi\omega_{\mathrm{p}}(1-\alpha_m^2)^{1/4}$$

where $eE_0 = \sqrt{2}m\omega_{\mathrm{p}}c(1-\alpha_m^2)^{-1/4}$ so that

$$\omega \approx \frac{\pi}{2}\left(\frac{\omega_{\mathrm{p}}^2mc}{eE_0}\right).$$

In both these limits one observes that pure longitudinal waves with frequencies less than ω_{p} can propagate in the plasma provided the electric fields associated with them are sufficiently high.

3.3 *Coupled longitudinal-transverse waves*

For this general case it is not possible to obtain exact solutions but some interesting analytical and numerical solutions have been obtained in limiting cases. It is also convenient in some cases to restrict oneself to plane polarized waves such that $\rho_y = 0$. One interesting limit, which was first pointed out by Akhiezer and Polovin [1] is that of $\beta \to \infty$. This corresponds to an almost transverse wave with $\rho_z \to 0$. In this case, (7, 9) can be simplified to

$$\frac{\mathrm{d}^2\rho_x}{\mathrm{d}\zeta^2}+\frac{1}{\beta^2}\frac{\rho_x}{(1+\rho^2)^{1/2}}=0, \tag{18}$$

$$\frac{\mathrm{d}^2\rho_z}{\mathrm{d}\zeta^2}+\frac{1}{\beta^2}\frac{\rho_z}{(1+\rho^2)^{1/2}}=0. \tag{19}$$

Equations (18, 19) are exactly integrable and yield periodic solutions with amplitudes oscillating between two limits which are determined by the initial values of the two constants of motion that the system permits. The other interesting limit is $\beta \to 1$, i.e., $\beta - 1 \ll 1$. In this case one can redefine the independent variable as $\eta = (\beta^2-1)^{-1/2}$ and rewrite (7—9) as

$$\frac{\mathrm{d}^2\rho_x}{\mathrm{d}\eta^2}+\frac{\rho_x}{(1+\rho^2)^{1/2}-\rho_z}=0, \tag{20}$$

$$\frac{\mathrm{d}^2\rho_y}{\mathrm{d}\eta^2}+\frac{\rho_y}{(1+\rho^2)^{1/2}-\rho_z}=0, \tag{21}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\eta^2}\left[\rho_z-(1+\rho^2)^{1/2}\right]+\frac{(\beta^2-1)\rho_z}{(1+\rho^2)^{1/2}-\rho_z}=0, \tag{22}$$

where $\beta - 1 \ll 1$ has been used. In (22) the last term can be neglected to yield $(1+\rho^2)^{1/2}-\rho_z = C^2$ where C is a constant. One then obtains the following periodic solutions

$$\rho_x = A_x\cos(\omega\eta); \quad \rho_y = A_y\sin(\omega\eta),$$

$$\rho_z = \frac{A_x^2-A_y^2}{4\sqrt{1+\frac{1}{2}(A_x^2+A_y^2)}}\cos(2\omega\eta),$$

where

$$\omega = \left[1+\tfrac{1}{2}(A_x^2+A_y^2)\right]^{-1/4}$$

and the amplitudes A_x and A_y are related to C by $A_x^2+A_y^2 = 2(C^4-1)$. Kaw and Dawson [4] carried out a detailed numerical investigation of the original equations (7, 9) and demonstrated the existence of periodic propagating solutions. They also confirmed that when $\beta \gg 1$ (phase velocity $\gg$ velocity of light), the waves are almost transverse. Furthermore, for strongly relativistic regime J vs ζ (where J is the current density) is in the nature of a square wave and correspondingly the electric field E as a function of ζ has a saw-tooth shape. Physically this is expected because in a strongly relativistic regime, $v_x \to c$ in most of the period and so $J_x \to -n_0ec$, giving the square wave form. One may now give a simple interpretation of propagation of strongly relativistic waves into overdense plasmas. Normally, the conduction current generated by the electron response to the wave fields acts as a shielding current giving the plasma a dielectric behaviour and preventing propagation of waves into overdense regions. Now since the shielding current saturates in the strongly relativistic regime, the waves can penetrate into overdense plasmas. Relativistic effects thus diminish the ability of the plasma to act as a dielectric.

An analytic description of the numerical results of Kaw and Dawson [4] was obtained by Max and Perkins [5] who also derived the propagation condition in the strongly relativistic regime as,

$$\frac{1}{\beta^2} = \frac{c^2k^2}{\omega^2} = 1 - \frac{\pi}{2}\frac{\omega_\mathrm{p}^2}{\omega}\frac{mc}{eE} > 0. \tag{23}$$

Max and Perkins [5] were also the first ones to extend the concept of these solutions to propagation into weakly inhomogeneous plasmas. They used a WKB like condition for constancy of energy flux for waves propagating into reflectionless inhomogeneous plasma and pointed out that indefinite penetration into overdense plasmas is possible if the following condition is satisfied,

$$\left(\frac{\omega L}{c}\right)^{1/2}\frac{eE_i\omega}{mc} > \omega_\mathrm{p}^2$$

where E_i denotes the electric field strength in a vacuum and L is the density scale length. Physically as the wave penetrates to a higher density, there is a WKB amplification of the electric fields which decreases the effective plasma frequency (due to increased effective mass) faster than its increase due to density changes. The above condition on L arises in ensuring that the WKB condition, viz. scale length of density variation $\gg$ the wavelength of the wave, continues to be satisfied indefinitely. The notion of indefinite wave propagation into overdense plasmas has some interesting physical applications in astrophysics particularly to radiation occuring in pulsars. Finally we would like to mention the article of DeCoster [6] in which a lot of the early work related to traveling wave solutions has been summarised. He has also carried out a weak coupling analysis and pointed out (without any amplification or discussion) the existence of nonlinear resonances whenever $2m = n(1 - 1/\beta^2)^{1/2}$.

3.4 *The nonlinear Hamiltonian*

For a more general overview of the linearly polarized traveling wave solutions we introduce the variables,

$$X = \rho_x(\beta^2 - 1); \quad Z = \beta\rho_z - (1 + \rho^2)^{1/2}$$

and rewrite (7, 9) in the following form

$$\ddot{X} + \frac{\beta}{\sqrt{\beta^2 - 1 + X^2 + Z^2}}X = 0, \tag{24}$$

$$\ddot{Z} + \frac{\beta}{\sqrt{\beta^2 - 1 + X^2 + Z^2}}Z + 1 = 0, \tag{25}$$

where the overdots refer to differentiation with respect to ζ. These equations can be derived from the Hamiltonian,

$$H = \frac{\dot{X}^2}{2} + \frac{\dot{Z}^2}{2} + \beta\sqrt{\beta^2 - 1 + X^2 + Z^2} + Z \tag{26}$$

which is a constant of motion. The coupled stationary problem is thus equivalent to that of the motion of a fictitious particle in a two dimensional potential. Note that for $\beta > 1$, the Hamiltonian is positive definite in the entire $X - Z$ plane and will yield bound solutions of X and Z. For $\beta < 1$, this property is destroyed and the possibility of unbounded runaway solutions exist [7].

The Hamiltonian (26) has one equilibrium point ($X = 0$, $Z = -1$) where $H = H_{\min} = \beta^2 - 1$. Physically this corresponds to $\rho_z = \rho_x = \dot{\rho}_z = \dot{\rho}_x = 0$, i.e., to the equilibrium plasma with no oscillatory motions. Linearisation of (24) and (25) around the equilibrium points, gives

$$\delta\ddot{X} + X = 0, \tag{27}$$

$$\delta\ddot{\tilde{z}} + \left(1 - \frac{1}{\beta^2}\right)\tilde{z} = 0, \tag{28}$$

where $\tilde{z} = Z + 1$. The 'frequency' $(1 - 1/\beta^2)^{1/2}$ corresponds to longitudinal oscillations at the plasma frequency, as can be verified by going back to the unnormalised variables. Similarly the 'unit' frequency of 'X' oscillation corresponds to linear electromagnetic waves. The limit $X = \dot{X} = 0$ corresponds to nonlinear longitudinal oscillations whose exact form has been discussed in an earlier section.

To explore the full range of nonlinear solutions, we carried out an extensive numerical investigation of (24, 25) using the Poincare surface of section technique. Basically, we solved for the 'particle orbits' of the Hamiltonian (26), and looked at the successive crossings of the orbits of the plane $X = 0$ (with $\dot{X} > 0$) for a given set of values of H and

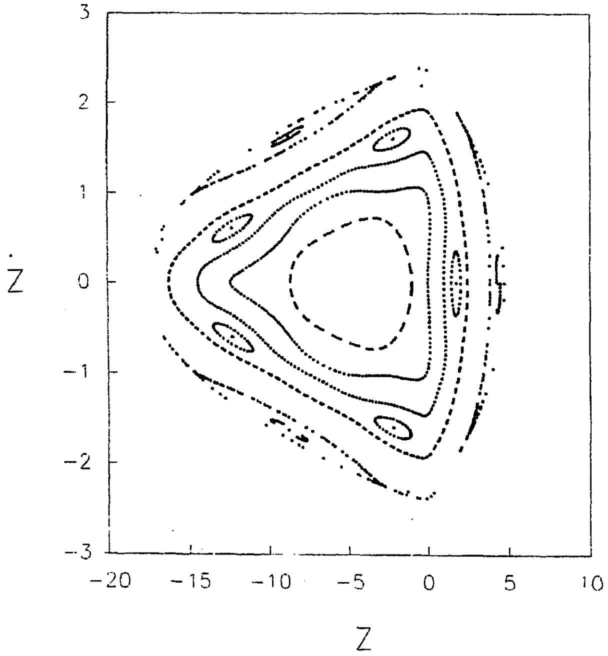


Figure 1. Poincare surface of section plot, $\dot{Z}$ versus Z ($X = 0, \dot{X} > 0$), for the Hamiltonian with $\beta = 1.1$, $\beta_b = 0.7$ and $H = 3.0$

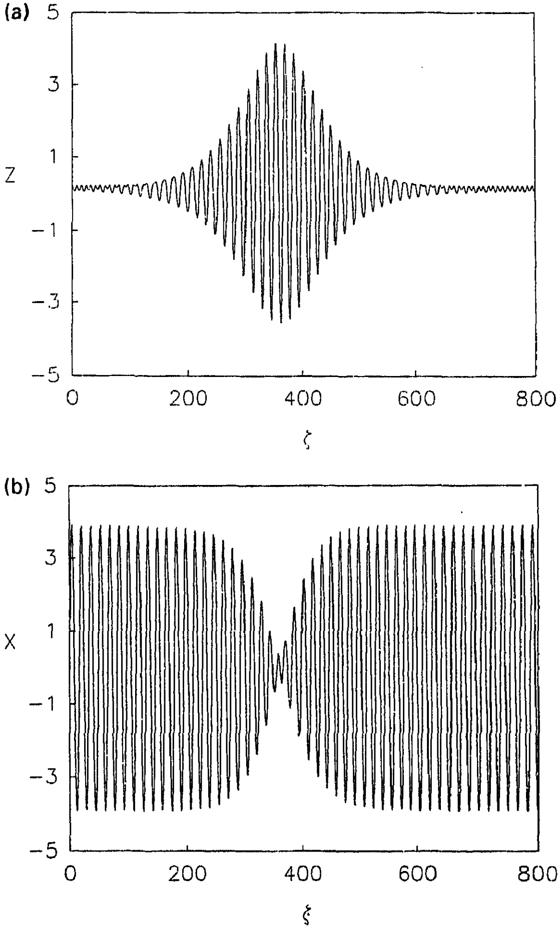


Figure 2. (a) A typical soliton solution of the Z oscillator corresponding to $H = 2, \beta = 1.3$ and $\beta_b = 0.8$, where the separatrix is between $(2.94, 0)$ and $(-3.305, 0)$. (b) A typical 'time' orbit curve of the X oscillator for the corresponding Z soliton.

β. Each orbit corresponds to a solution of the coupled equations. In general we saw the following kind of orbits: (i) Fixed points which map back on themselves after a finite number of iterations. These correspond to periodic orbits in which the ratio of the frequencies of the two oscillators is a rational number. (ii) Quasi-periodic orbits which ergodically fill a curve in the $Z - \dot{Z}$ plane and for which the ratio of the frequencies is an irrational number. (iii) Islands surrounding a fixed point which correspond to amplitude modulated waves. (iv) Separatrix orbits between chains of islands which correspond to amplitude modulated waves with the modulation period being infinite. Physically these correspond to soliton solutions. Another possibility which exists in principle (but which we did not find in our early numerical work) is stochastic orbits for which the orbit ergodically fills an area in space. These would correspond to aperiodic orbits which are

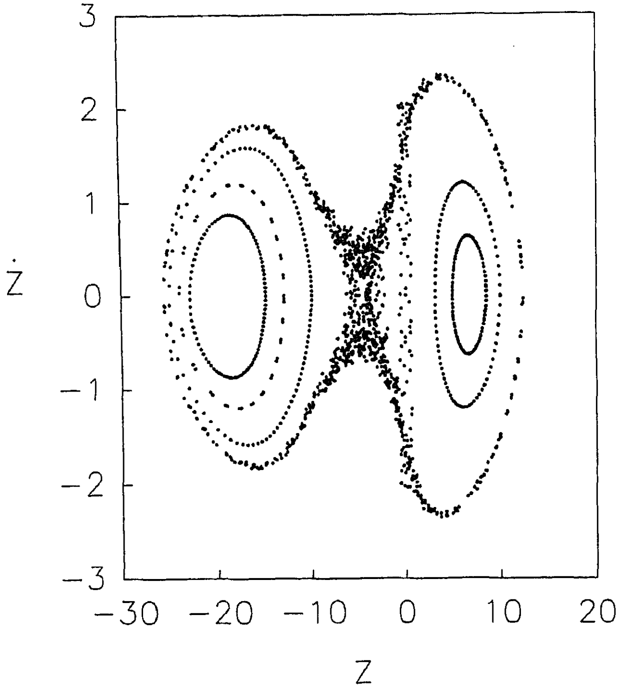


Figure 3. Chaotic orbits in Poincare surface of section plot, $\dot{Z}$ versus Z ($X = 0, \dot{X} > 0$), for the Hamiltonian with $\beta = 1.21$, $\beta_b = 0.7$ and $H = 10.0$

highly sensitive to initial conditions. The absence of numerical evidence for stochastic orbits and the results of local linear analysis of the trajectories led us to conjecture that perhaps the Hamiltonian was integrable. The analytic structure of the Hamiltonian, particularly the presence of the squareroot term in the potential, posed a challenge for application of standard integrability tests such as the Painleve criterion. There was thus widespread interest for a while to provide an analytic proof of its integrability. It came as somewhat of a surprise when Grammaticos, Ramani and Yoshida [3] proved that it was nonintegrable. Their proof was based on a clever application of Ziglin's theorem. Subsequently numerical evidence of stochastic orbits was also found by Romeiras [8]. In a recent work, Bisai, Sen and Jain [9] have carried out further detailed studies on a more generalised form of the Hamiltonian (26), namely

$$H = \frac{\dot{X}^2}{2} + \frac{\dot{Z}^2}{2} + \mid \beta - \beta_b \mid \sqrt{\beta^2 - 1 + X^2 + Z^2} + (1 - \beta\beta_b)Z \tag{29}$$

In this the plasma is assumed to have an equilibrium velocity v_b (beam velocity) which is represented by the normalised quantity β_b. Figure 1 shows a typical Poincare plot of orbits obtained from (29) and displays the different kind of solutions including the 'island' orbits. Figures (2a, b) display the 'soliton' orbits corresponding to the separatrices on the Poincare plots. In figure 3 we see the presence of chaotic orbits in the Poincare plot. Note that for $\beta = 1/\beta_b$, the Hamiltonian is integrable, since it admits $L \equiv Z\dot{X} - X\dot{Z}$ as another invariant (which is in involution to H, i.e., $[H, L] = 0$). This is solutions on the Poincare surface) to provide a detailed theoretical basis for the earlier conjectures of DeCoster [6] regarding the nonlinear resonances.

4. Relativistically intense standing waves

Marburger and Tooper [10] and Lai [11] found a new class of solutions to the Akhiezer–Polovin equations (1–6), which correspond to relativistically intense standing waves in an overdense plasma. They introduced a new ansatz, namely, $\mathbf{E}(z,t) = \mathbf{F}(z)f(t)$ and restricted themselves to circular polarizations only. The conservation of canonical momentum in the transverse direction then led to the condition $\mathbf{p}_\perp + e\mathbf{A}/c = 0$ and the longitudinal momentum p_z was found to vanish. Here the vector potential $\mathbf{A}$ follows the ansatz

$$\frac{e\mathbf{A}}{mc^2} = F(\zeta)(-\hat{x}\cos(\omega t) + \hat{y}\sin(\omega t)),$$

where $\zeta = \omega z/c$ and F satisfies the equation

$$\frac{\mathrm{d}^2F}{\mathrm{d}\zeta^2} + F = \left(\frac{\omega_\mathrm{p}^2}{\omega^2}\right)\frac{F}{(1+F^2)^{1/2}}. \tag{30}$$

The electron density n and hence ω_p^2 is bunched in the longitudinal direction because the standing wave produces ponderomotive forces which push the electrons away from the field maxima. Since ions are assumed to be frozen on the time scales of interest, the electrons are held in balance by generation of static longitudinal fields $\Phi(z) = (mc^2/e)\phi(\zeta)$ where ϕ satisfies the equation

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}\zeta^2} = \left(\frac{\omega_{\mathrm{p}0}^2 - \omega_\mathrm{p}^2}{\omega^2}\right) = \frac{\mathrm{d}^2}{\mathrm{d}\zeta^2}(1+F^2)^{1/2}. \tag{31}$$

One may now eliminate ω_p^2 from (30) and reduce it to that of a nonlinear one dimensional oscillator governed by the Hamiltonian

$$H = \frac{P^2}{2M(F)} + \tfrac{1}{2}F^2 - \alpha(1+F^2)^{1/2} \tag{32}$$

where $P = M(F)(\mathrm{d}F/\mathrm{d}\zeta), M(F) = (1+F^2)^{-1}$ and $\alpha = \omega_{\mathrm{p}0}^2/\omega^2$. For overdense plasmas ($\alpha > 1$), this leads to three types of nonlinear standing waves, viz,

(1) oscillations about zero when the maximum field F_m exceeds a critical value $F_T = 2(\alpha^2 - \alpha)^{1/2}$;
(2) oscillations about a bias field $F_B = (\alpha^2 - 1)^{1/2}$ when $F_m < F_T$ and
(3) motion with an infinite period (along the separatrix) when $F_m = F_T$.

For $F_m < F_a = 2(\alpha^2 - \alpha)^{1/2}$, the equations for F may be integrated explicitly in terms of Jacobian elliptic functions. Thus, when $F_T \leq F_m \leq F_a$, the quartic period is $\zeta_\mathrm{p} = K(m)/\gamma$

where K is the complete elliptic integral of the first kind, $m = (\epsilon_m - 1)(2\alpha + 1 - \epsilon_m)/4(\epsilon_m - \alpha)$, $\gamma^2 = \epsilon_m - \alpha$, $\epsilon_m^2 = 1 + F_m^2$ and the field amplitude is given by

$$F(\zeta) = \frac{2F_m cn(\gamma\zeta)}{[2 + (\epsilon_m - 1)sn^2(\gamma\zeta)]}.$$

For the important nonperiodic limit, $F_m = F_T$, $m = 1$ and

$$F(\zeta) = \frac{F_T \cosh(\gamma''\zeta)}{[1 + \alpha \sinh^2(\gamma''\zeta)]},$$

where $\gamma'' = (\alpha - 1)^{1/2}$.

An important physical effect which has to be carefully considered is the possibility of complete depletion of electrons from a part of the region occupied by the standing waves. The condition for complete vanishing of the electron density may be written down as

$$\frac{\mathrm{d}^2}{\mathrm{d}\zeta^2}(1 + F^2)^{1/2} < -\alpha.$$

If this condition is satisfied somewhere, part of the standing wave is in a vaccuum cavity; appropriate solutions are then obtained inside the cavity and their logarithmic derivatives are matched across the depletion boundary. One can show that for $F_m < F_d = \{(\alpha/2)[\alpha + (4 + \alpha^2)^{1/2}]\}^{1/2}$ complete electron depletion does not take place anywhere and the above simple analytic solutions are valid. Elsewhere, the solutions have to be obtained with greater care.

The most interesting application of these solutions is in estimating the properties of standing waves in inhomogeneous plasmas. Using the WKB condition that the phase integral $I = \oint P\mathrm{d}F$ is an invariant for propagation into an inhomogeneous plasma with $\alpha = \alpha(\zeta)$, one finds the condition

$$I_0 = \pi F_0^2 = I_T = 8\alpha \sin^{-1}[(\alpha - 1)/\alpha]^{1/2} - 8(\alpha - 1)^{1/2}.$$

Here F_0 is the vacuum field strength and I_T, the phase integral corresponding to the turning point, is obtained by using the solution for the case $F_m = F_T$ (since in this case the period $\zeta_\mathrm{p} \to \infty$). Note that in contrast to the linear WKB solution, the wave amplitude remains finite at the turning point: $F_T = 2(\alpha^2 - \alpha)^{1/2}$. The above condition may be used to find out the maximum density upto which an incident wave with field amplitude F_0 may penetrate.

5. Modulated nonlinear waves

In this section, we will discuss another class of exact one dimensional nonlinear solutions of the relativistic cold plasma equations which represent envelope solitons of light waves [12, 13], in which the modulation envelope propagates as a large amplitude plasma wave in the medium. These solutions are a step beyond the well known traveling wave solutions in a cold plasma because the envelope and the phase propagate with different speeds so that the nonlinear relationships between the phase and group velocities can be investigated. For simplicity we again restrict our analysis to circularly polarized waves which couple to longitudinal disturbances only because the amplitude is modulated.

We start as usual with the relativistic set of fluid equations (1–6 for a cold plasma (in one dimension) and introduce the change of variables $x-\beta t=\xi$, $t=\tau$. Instead of the fluid momenta (ρ) we now work with A (the vector potential) and ϕ (the electrostatic potential) which are normalised as eA/mc^2, $e\phi/mc^2$. Time is normalised as $\omega_p t$ and space by $\omega_p x/c$, where the direction of propagation is taken as x. We now make the key ansatz that in the moving frame, the vector potential is circularly polarized and has a sinusoidal phase variation, i.e., $\mathbf{A}=[\{\hat{y}a(\xi)+\hat{z}ia(\xi)\}\exp(-i\lambda\tau)+\text{ c.c}]$. The introduction of a frequency parameter λ in the phase factor is the basic deviation from earlier stationary wave solutions; this allows us to identify β with the group velocity of the light pulse and also to distinguish between the group and phase speeds. The choice of circular polarization allows us to avoid the generation of harmonics in all wave fields. For plasma oscillations, which form the modulation envelope, this leads to $\partial/\partial\tau=0$ in the moving frame. Integrating the fluid equations we then get $n(\beta-u)=\beta$ and $\gamma(1-\beta u)-\phi=1$. Using these relations and further writing $a(\xi)=R\exp(i\theta)$, we can reduce the Poisson's equation and the wave equation to obtain the following set of coupled nonlinear equations

$$\phi''=\frac{u}{(\beta-u)} \tag{33}$$

with u given by

$$u=\frac{\beta(1+R^2)-(1+\phi)[(1+\phi)^2-(1-\beta^2)(1+R^2)]^{1/2}}{(1+\phi)^2+\beta^2(1+R^2)} \tag{34}$$

and

$$R''+\frac{R}{1-\beta^2}\left[\left(\lambda^2-\frac{M^2}{R^4}\right)\frac{1}{1-\beta^2}-\frac{\beta}{\beta-u}\frac{1-\beta u}{1+\phi}\right]=0, \tag{35}$$

where $M=R^2\{(1-\beta^2)\theta'-\lambda\beta\}$, is a constant of integration. In our representation the amplitude $R=(A_y^2+A_z^2)^{1/2}$ of the circularly polarized electromagnetic wave exhibits a modulation propagating at the group speed β. The question of the definition of phase speed is a little more complex. For small amplitudes R is always positive and the exponential factor contains the entire phase information. The phase speed may then be shown to be simply $1/\beta$ satisfying the conventional relation $V_pV_g=1$. However, for arbitrary amplitudes R oscillates because of the strong $R-\phi$ coupling and the phase speed has to be defined by directly determining the z,t variation of $R\cos(\theta-\lambda\tau)$. Equations (33–35) admit of one exact integral of motion which may be written as

$$K=\frac{R'^2}{2}-\frac{\phi'^2}{2(1-\beta^2)}+V(R,\phi), \tag{36}$$

where

$$\begin{aligned}V(R,\phi)=&\frac{\lambda^2}{(1-\beta^2)^2}\frac{R^2}{2}+\frac{M^2}{2R^2(1-\beta^2)^2}-\frac{\phi}{(1-\beta^2)}\\&-\frac{\beta}{(1-\beta^2)^2}[\beta(1+\phi)-\{(1+\phi)^2-(1-\beta^2)(1+R^2)\}^{1/2}].\end{aligned} \tag{37}$$

Since $\beta < 1$ (group velocity less than c), the problem is similar to a Hamiltonian of coupled anharmonic oscillators with two degrees of freedom (R, ϕ) where the effective mass for one of the anharmonic oscillators is negative. In the limit of weak density response, the problem can be made one dimensional by the substitutions $\phi \approx (1+R^2)^{1/2} - 1$ and $n \approx (\sqrt{1+R^2})''$. Expanding in R^2 and taking $M = 0$, $K = 0$ we get the well known envelope soliton solution

$$A_y = R_m \operatorname{sech}\left[\frac{\sqrt{1-\beta^2-\lambda^2}}{1-\beta^2}(x-\beta t)\right] \cos\left[\frac{\lambda\beta}{1-\beta^2}\left(x-\frac{t}{\beta}\right)\right], \tag{38}$$

where

$$R_m = 4\sqrt{\frac{(1-\beta^2)(1-\beta^2-\lambda^2)}{4\lambda^2-(1-\beta^2)(3+\beta^2)}}. \tag{39}$$

$K \neq 0$ gives modulated periodic wave train solutions with the envelope factor described by an elliptic function. Equation (38) represents an isolated light pulse with a frequency $\omega = \lambda/(1-\beta^2)$, wavenumber $k = \lambda\beta/(1-\beta^2)$, phase velocity $1/\beta$, group velocity β (note $V_p V_g = c^2$), an envelope scale length (in units of c/ω_p) equal to $(1-\beta^2)/\sqrt{(1-\beta^2-\lambda^2)}$ and an amplitude-group velocity relationship given by (40). Note that the envelope scale length is real only when $\lambda^2 < (1-\beta^2)$. This may be physically interpreted as follows. The wave-frequency in the frame moving with the group velocity has the Doppler shifted value $\bar{\omega} = (\omega - k\beta)/\sqrt{1-\beta^2} = \lambda/\sqrt{1-\beta^2}$. Thus the inequality $\lambda < \sqrt{1-\beta^2}$ corresponds to the situation when the Doppler shifted wave frequency $\bar{\omega}/\omega_p < 1$, i.e. the electromagnetic wave finds itself in an overdense plasma and is totally trapped; this is why we get a soliton solution and the wave does not leak out. The effective scale-length for trapping may now be seen to be $\sqrt{1-\beta^2}[c/\sqrt{\omega_p^2 - \bar{\omega}^2}]$ which is quite reasonable. Similarly the amplitude-group velocity relationship may be written in terms of physical parameters as

$$1-\beta^2 = \frac{-(R_m^2 - 4 + 4R_m^2\omega^2) + \sqrt{((R_m^2 - 4 + 4R_m^2\omega^2)^2 + 64R_m^2\omega^2)}}{8\omega^2}. \tag{40}$$

For arbitrary amplitudes we have numerically solved equations (33–35) for soliton solutions. For λ very close to $\sqrt{1-\beta^2}$ we get small amplitude solitons which are well described by the analytic solutions (38). As λ is decreased the soliton amplitude increases and acceptable solutions occur only at discrete values of λ. In other words, for a fixed value of β finding soliton solutions turns out to be an eigenvalue problem in λ. The sizes and shapes of these solitons also vary as a function of λ. Typically ϕ has a characteristic bell shape whereas R has a number of nodes. Figure 4 shows a typical soliton solution for $\beta = 0.97$ and $\lambda = 0.224445$. For applications such as particle acceleration or photon acceleration, the regime of interest is $\beta \rightarrow 1$, where the group velocity is close to c. We have carried out a detailed investigation of soliton solutions in this regime. Figure 5 is a plot of normalized group velocity V_g versus the normalized carrier frequency $\Omega(= \omega/\omega_p)$. The solid curve corresponds to soliton pulse results obtained from our numerical work. For comparison we have also plotted the linear group velocity (dashed line)

$V_{gL} = \sqrt{(1 - 1/\Omega^2)}$ and the nonlinear group velocity for the infinite plane wave (dotted line), $V_{g\infty} = \beta[1 - (\gamma_\infty - 1)/(2\Omega^2\gamma_\infty(\gamma_\infty + 1)]$. This last expression has been recently obtained by Mori *et al* [14] and depends on the infinite wave γ viz. $\gamma_\infty = \sqrt{1 + A^2}$. The nonlinear group velocities are closer to c than the linear group velocity. This is physically understandable because the nonlinearity makes the electrons heavier and thereby weakens the plasma dielectric effects. Finite width soliton pulses propagate slower than the infinite plane waves. Physically this is because coupling to plasma waves acts as a drag on the electromagnetic waves and slows them down. The deviations between $V_{g\infty}$ and V_{gs} can be substantial; in our numerical work, we have observed (not shown) deviations upto 25 per cent of the difference between V_g and 1.

To understand these solutions better we consider another approximate analytic limit to equations (33–35) particularly applicable when the ϕ soliton structure is large. When $\phi_{\max}$ is large, the longitudinal velocity $u \to -1$, $n \to \beta(1 + \beta)^{-1}$ and $n/\gamma \to 0$. In this case, the light wave propagates essentially in a plasma free region (because the residual electrons have become infinitely massive) and we can approximate the effective Hamiltonian as

$$K_{\text{eff}} = \frac{R'^2}{2} + \frac{\lambda^2}{(1-\beta^2)^2}\frac{R^2}{2} - \frac{\phi'^2}{2(1-\beta^2)} - \frac{\phi}{(1-\beta^2)(1+\beta)} + \frac{\beta}{(1+\beta)(1-\beta^2)} + \cdots, \tag{41}$$

where the $\cdots$ refers to terms of order $\leq \phi^{-1}$ which are neglected. From (41) we get $R = R_0 \sin(\lambda\xi/(1 - \beta^2))$. Furthermore, the structure of ϕ near its maximum can be

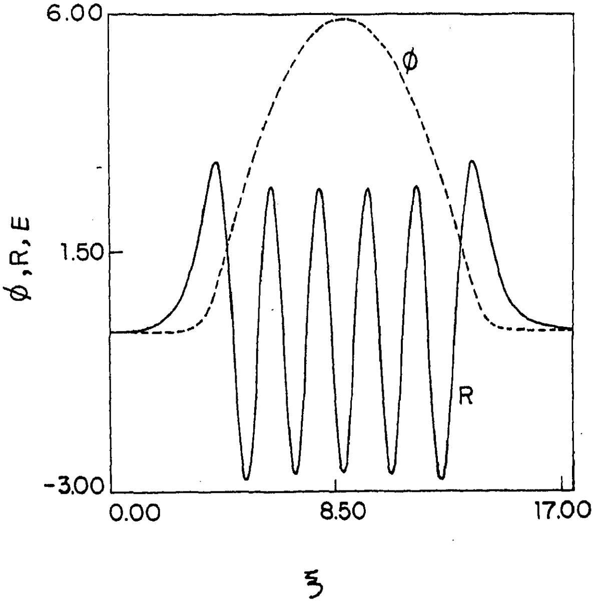


Figure 4. Electrostatic potential ϕ (dashed line), and electromagnetic wave amplitude (R) (solid line) profiles for a soliton pulse with $p = 6$, $\beta = 0.97$ and $\lambda = 0.224445$.

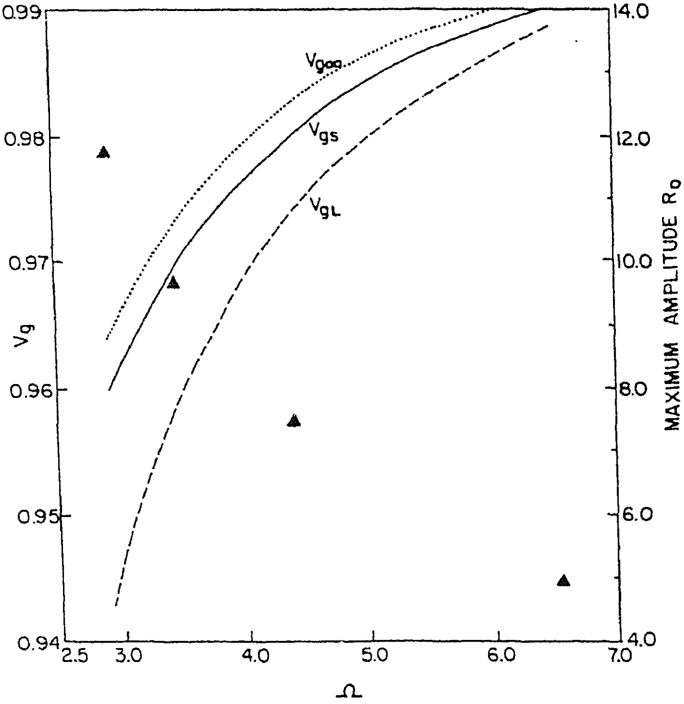


Figure 5. Normalized group velocity vs normalized frequency curves for soliton pulses (solid line), infinite pulses (dotted line) and linear pulses (dashed line). The maximum electromagnetic wave amplitude R_0 for the soliton pulse (for $p = 18$) are indicated with solid triangles.

approximated by the parabolic equation ($u \to -1$ in (33)) $\phi = \phi_{\max} - \xi^2/2(1+\beta)$, which gives an estimate of the spatial scale size of the soliton as $\xi_{\max} \simeq \sqrt{2(1+\beta)\phi_{\max}}$. From (36, 37) we know that $K = 0$ for a soliton solution. Noting that $K \simeq K_{\text{eff}}$ for $\phi = \phi_{\max}$, $\phi' = 0$, we get $\phi_{\max} = \beta + \lambda^2 R_0^2/2(1-\beta)$. Furthermore, if the soliton region involves p wavelengths of the light wave, we have $p\pi(1-\beta^2)/\lambda \simeq \xi_{\max}$. The above expressions lead to the following approximate analytic relationship between λ, R_0 and β as

$$R_0^2 = \frac{2(1-\beta)}{\lambda^2}\left[\frac{\pi^2 p^2}{2\lambda^2}(1-\beta^2)(1-\beta) - \beta\right]. \tag{42}$$

Equation (42) is largely corroborated by the numerical work shown in figures 4 and 5.

Recently, Kuehl *et al* [15] have carried out a WKB analysis in the weak amplitude limit of eqs (33–35) and have obtained a more refined estimate of the discrete spectrum. They have also shown that the discrete nature of the eigenvalue spectrum persists as long as the dispersion term in the Poisson equation is retained. If one neglects the dispersion then $\phi \propto R^2$ in the small amplitude limit and substituting into the R equation leads to the nonlinear Schroedinger equation which has a continous spectrum of solitons. Thus the coupled electromagnetic-plasma wave solitons have a unique characteristic. The importance of the dispersive term in the formation of these solitons has also been pointed recent work of Decker and Mori *et al* [17] who have carried out particle simulations for this purpose.

6. Discussion

We have discussed the nonlinear propagation of relativistically intense electromagnetic waves into collisionless overdense plasmas putting a special emphasis on 1-d plane wave solutions of the propagating, standing and modulated (i.e. localized) types. These solutions already exhibit a rich variety of phenomena associated with relativistic electron mass variations and coupling between transverse electromagnetic and longitudinal fields. However, it must be admitted that whereas these special solutions are excellent candidates for elucidating the physics of the important nonlinear phenomena, their relevance to truly experimental situations may be determined by questions related to their accessibility. Thus one may need very special initial conditions for setting them up and one may have to worry about their stability to noise perturbations in one and more dimensions. Many of these questions have not yet been addressed in much detail.

One question which has received some attention is that of self-focusing and self-modulational instability of relativistically intense electromagnetic wave [18, 19]. Self-focusing instability occurs when the field power exceeds a critical value and leads to filamentation of intense electromagnetic waves [20, 18] Similarly a self-modulational instability could be responsible for converting a propagating plane wave solution (§ 3) into one with localized modulational envelopes (§ 5). If both instabilities are simultaneously occuring, the final object would be a multi-dimensional localized region ('bubble') of intense electromagnetic waves interacting with longitudinal fields as they propagate. The analytic description of such 'bubbles' still remains an outstanding problem in nonlinear plasma physics. The strong low frequency electromagnetic radiation believed to be present in the pulsar environments [21, 22] is the most important astrophysical problem where these results have been applied. Retaining relativistic effects in the ion motions also, it has been concluded that the pulsar radiation can penetrate to large distances in the Crab nebula, thus allowing the possibility of an energy source in these volumes [5].

In laboratory plasmas, the most important application of this work is in laser fusion and plasma based accelerators. In the inertial fusion schemes, a point of view has recently developed that it would be of interest to create a hot spot in the central core of a compressed inertial fusion pellet by some independent means [23]. The use of a relativistically intense light beam might be one possible method of reaching such overdense regions. Similarly, for plasma based accelerator schemes the creation of space charge fields propagating with velocities close to that of light is of considerable interest [24–26]. A proper nonlinear description of the group velocity of such coupled stationary waves is therefore under intense investigation [13, 15, 14]. The solitary light pulses which are mostly empty of plasma in the central region, but have large plasma accumulation at the edges, have consequently large changes in n/γ inside the pulse. One

In this paper, we are concerned with the demonstration that the universal properties of the quasiperiodic route to chaos in the sine circle map ([2], [3], [4]) is an appropriate framework for describing the low-Reynolds-number behavior of the wake behind an oscillating cylinder. Previous experimental studies of this nature in this particular flow can be found in Refs. [5] and [6], and similar studies in other fluid flows in Refs. [7], [8], [9], and [10]. This paper is in some respects a synopsis of results obtained a few years ago [11], but it also presents some that have not been published before.

2. The flow

We study the wake of an oscillating circular cylinder of diameter 0.5 mm, housed in a low-noise suction-type wind tunnel, and oscillated electromagnetically in the direction transverse to that of upstream fluid velocity. In practice, the oscillating cylinder is a stainless steel string of uniform diameter stretched taut, through the wind tunnel walls, over a distance that is four times the tunnel span. The active length/diameter ratio of the cylinder is about 300. The Reynolds number based on the cylinder diameter and the velocity of the oncoming air flow is 55. We place a velocity sensor at a distance of 15 diameters behind the cylinder, half a diameter off the geometric axis of the stationary wake. The velocity sensor is a standard hotwire of 5 μm diameter and 0.6 mm length, operated on a constant temperature anemometer. The sensor measures the streamwise component of fluid velocity. For more details of the experimental arrangement, instrumentation and data processing, please see Ref. [11].

The regular 'shedding' of vortices at Reynolds numbers above about 40 imparts to the wake a periodic motion of frequency f_o, say. The frequency f_o is a known function of the Reynolds number [12]. The transverse oscillation at the controlled frequency f_e imposes a second frequency on the wake. The amplitude of oscillation of the cylinder can be thought of as a coupling parameter between vortex shedding and cylinder oscillation. This nonlinear coupling "shifts" the shedding frequency to some neighboring value, say f_o', which depends on the ratio $\Omega = f_e/f_o$ and the magnitude of the nonlinear coupling between the two oscillators. In the jargon of quasiperiodic dynamical systems, the parameter $\Omega = f_e/f_o$ is called the bare winding number, and the ratio $\omega = f_e/f_o'$ is called the dressed winding number.

3. The circle map

Systems with two independent frequencies f_o and f_e define a torus in the phase space. As the system evolves, the phase space trajectory hits a Poincaré section at some discrete iterates along a circle. The angular position of the iterates is the variable of dynamical interest. A paradigm situation is given by the sine circle map

$$\theta_{n+1} = \theta_n + \Omega - \frac{K}{2\pi}\sin(2\pi\theta_n), \tag{1}$$

where θ_n is the angular coordinate of the n^{th} iterate, Ω is the bare winding number and $K > 0$ is the nonlinear coupling parameter. The map is utilized by setting $0 < \Omega < 1$ and obtaining the iterates of θ. The dressed winding number is defined as $\omega = \lim_{n\to\infty}[(\theta_n - \theta_o)/n]$.

The experimental situation is qualitatively similar to that encountered in circle maps; following the spirit of universality mentioned above, it appears reasonable to explore the analogies quantitatively. For this purpose, we should recall a few key results about the circle map, known from the theoretical analysis of Refs. [2]–[4]:

- For low forcing amplitudes, equation (1) exhibits Arnol′d tongues (regions of mode locking) in which the frequency f_o shifts so that the dressed winding number is fixed as a rational over some finite range of f_e.
- In other case where the dressed winding number is irrational, the motion is quasiperiodic.
- As the nonlinear coupling parameter K increases, the Arnol'd tongues widen; they overlap at some critical line corresponding to the critical value of K. This critical line corresponds to the breakdown of the two torus and the onset of chaos.
- Along the critical line, the system exhibits the devil's staircase structure.
- A special case of interest is the locus of the position in the $K - \Omega$ space where the dressed winding number is the inverse of the golden mean. (Loosely speaking, golden mean $\equiv (\sqrt{5} - 1)/2$ is the "most" irrational number, or least well represented by a rational approximation, because its continued fraction representation contains only the numeral one.) At the critical point on this line, where $\omega = \sigma_g$, the latter being the inverse of the gloden mean, the system is least influenced by mode-locking and is quasiperiodic. The properties of the system at this critcal point are universal.
- At the critical point, the power spectrum of the angular variable has a certain well-quantified self-similar structure [2].
- A description of this behavior requires modern dynamical systems theory including multifractals [13] and scaling functions [2].

4. Devil's staircase and power spectral density

In Refs. [5] and [11], it was shown that most of these properties were exhibited by the wake of the oscillating cylinder. As a quick recapitulation, we show in figure 1 the devil's staircase structure along the critical line. The small circles mark the extremes of a few principal Arnol′d tongues corresponding to the rational values of the dressed winding number marked on the figure. These are to be compared with the horizontal bars representing the theoretical results for the circle map. The comparison is good on the whole. Many other tongues were mapped out experimentally but are not given on the figure. A few narrower tongues near the rational numbers $4/15$ to $5/16$ are displayed in the inset. The general pattern is the same as that of the circle map, although some visible differences occur.

As a second example of the correspondence between the wake and the circle map, we show in figure 2 the scaled power spectral density of the velocity signal obtained close to the critical point. As known for the circle map dynamics, we have (nearly) equal-amplitude spectral peaks, marked 1, at powers of σ_g up to the fifth. For higher powers, the flow system does not respond like the circle map, yet the appearance of these equally-spaced peaks is quite encouraging. Furthermore, between any two of the four right-most

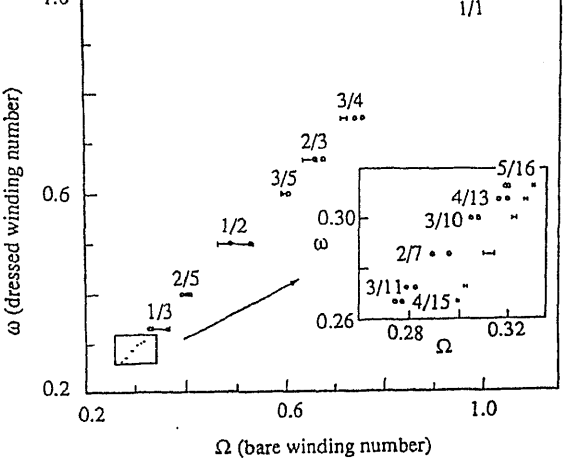


Figure 1. The devil's staircase showing mode-lockings along an experimentally determined critical line (circles). The general pattern is similar to that of the circle map (bars). The inset expands the small rectangle to the lower left.

primary spectral peaks, there are families of peaks marked 2, 3, 4 and 5, which—where they exist—appear at the right place according to the circle map theory. More delicate fine structure in the circle map is not replicated by the fluid system. It is difficult to say if these differences are real or if the fluid system is not exactly at the critical point as the theory demands. Locating the critical point is quite delicate, and the results are conspicuously sensitive to the precise location in its neighborhood [11]. Independent of this uncertainty, it is fair to say that the wake behaves, at least grossly, much like a circle map. This then suggests that the underlying dynamics of this fluid system may live on a low-dimensional attractor whose Poincaré section closely resembles that generated by the circle map. This expectation is explored in greater detail below.

5. Multifractal analysis

To construct the experimental Poincaré section at the critical point in phase space, the continuous hotwire velocity time trace $u(t)$ is sampled discretely at a time period τ equal to the period of excitation, yielding $P_n = u(t + n * \tau)$, $n = 0, 1, 2, 3, \ldots$. This procedure can lead to some problems if there is a drift in the bare or dressed winding numbers. A better procedure is to select points in the continuous time trace which have zero time derivative. In either case, the P_n are embedded in the three-dimensional phase space (P_n, P_{n+1}, P_{n+2}), three being the lowest dimension in which the P_n are non-interacting. A two-dimensional projection of the three-dimensional Poincaré section is shown in figure 3. The attractor lies essentially on a one-dimensional subspace and is perceptibly inhomogeneous. (Other projections described in Ref. [11] confirm this conclusion.)

We first turn to a suitable statistical characterization of the attractor. From the theory of circle maps, one knows that the attractor generated from the sine circle map has a

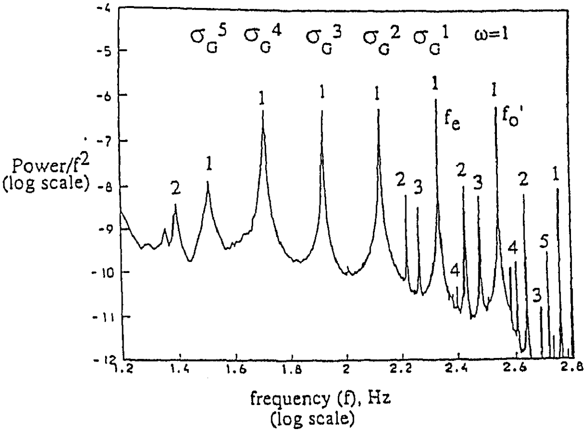


Figure 2. The scaled power spectral density for the velocity signal measured in the wake of the oscillating cylinder. The dressed winding number is estimated to be within 0.1 % of σ_g. The structure is quantitatively similar to that of the circle map.

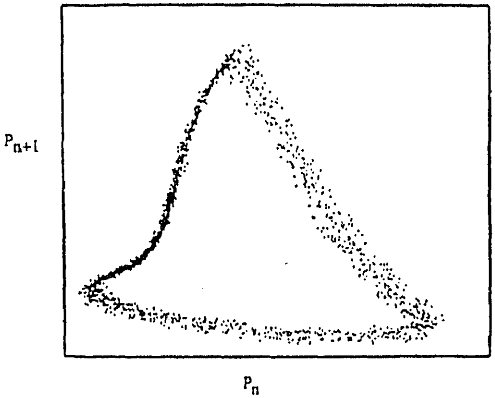


Figure 3. Two-dimensional projection of the experimental Poincaré section at the critical golden mean point, obtained by embedding the velocity signal in three dimensions as described in the text.

particular self-similar structure that can be quantified *via* the multifractal formalism. Exploitation of this self-similar structure for making predictions about the dynamics of the system is the goal of the multifractal formalism. To this end, one can construct a measure on this attractor by subdividing it into boxes of equal size and computing the density of points in each of these boxes. This gives one a measure of how often certain regions of the attractor are visited in comparison with others. One defines a singularity strength [13] by

$$\pi_i \sim L^{\alpha_i}, \tag{2}$$

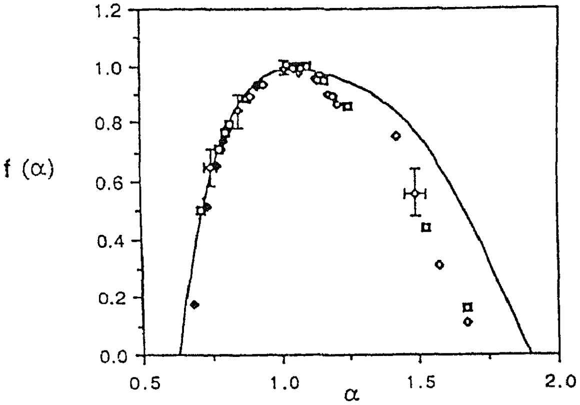


Figure 4.. The $f(\alpha)$ curve obtained from analyzing the measure on the Poincaré section. The solid line is the theoretical curve for the sine circle map at the golden mean critical point. Typical error bars are shown.

where π_i is the relative density of points in the i^{th} box of size L. The box size L is restricted to the scaling bounded by some inner and outer cut-off scales. Within the scaling range, one can define $f(\alpha)$ as the dimension of the set of points possessing singularity strength α by the equation

$$N(\alpha) \sim L^{-f(\alpha)}, \tag{3}$$

where $N(\alpha)$ is the number of boxes of size L possessing singularity strength α. The $f(\alpha)$ formalism entails viewing the attractor as the union of several interwoven fractal sets each having a measure of a different singularity strength. (In equations (2) and (3), the symbol $\sim$ subsumes all the prefactors needed to render them dimensionally homogeneous.)

A summary of these multifractal results is given in figure 4 which shows a comparison between the measured $f(\alpha)$ curve at the experimentally determined critical point and that for the sine circle map. Space limitations preclude a description here of how the curve was obtained from experimental data. This information can be found in Ref. [11]. On the left part of the figure representing the scaling of high probability regions on the attractor, the data agree well with the circle map predictions. The right half of the $f(\alpha)$ curve represents the scaling of the low-probability events on the attractor and is usually unreliable in experiment because of the strong influence of noise (even if small). It is thus unclear at present whether the discrepancy on the right part represents a true departure from the universal scaling or is an experimental artifact.

We have discussed this issue in more detail in Refs. [11] and [14] where we have concluded that the correspondence between the circle map and the flow is indeed real.

6. The trajectory scaling function

One would like to push further the correspondence between the flow and the circle map. One powerful way of describing the dynamics of a system is the so-called Feigenbaum scaling function [15], [16]. This object contains microscopic information about a deterministic dynamical system and its scaling properties. At the onset of chaos, the attractor can be regarded as constructed by an underlying multiplicative process that successively refines it, in such a way that it possesses scaling properties at finer and finer scales. The scaling function gives a compact description of the refinement basic to the fractal structure of the attractor. The underlying multiplicative process can be mapped on to a subdividing tree structure where, on the average, each branch divides into 2 subbranches at each successive level of refinement. The iterates of the dynamical system become branches of the tree, and intervals are formed by joining nearest neighbor branches.

At the critical golden mean winding number, the attractor can be thought to have been generated by successive refinements of a measure created from superstable periodic orbits whose winding numbers are successive Fibonacci approximations to the golden mean. The Fibonacci sequence is given by the relation $F_{i+2} = F_i + F_{i+1}$ with $F_o = 0$ and $F_1 = 1$. It is easily verified that the ratio $p/q = F_i/F_{i+1}$ approaches σ_g for large enough i. Since the intervals at each level are formed by joining nearest neighbor points, a measure is constructed to give equal weight to each such interval. The scaling function, which describes the contraction factors of each interval along each branch as the tree subdivides, is defined as

$$\sigma_\ell(\epsilon_{n+1}, \ldots, \epsilon_o) = \frac{L(\epsilon_{n+1}, \ldots, \epsilon_o)}{(\epsilon'_n, \ldots, \epsilon'_1, \epsilon_o)} \delta_{\epsilon_n, \epsilon'_n, \ldots, \epsilon_1, \epsilon_1^{ime}}, \tag{4}$$

where $L(\epsilon_n, \ldots, \epsilon_o)$ is the length of an interval belonging to the tree. The Kronecker delta's ensure that the two intervals being compared have the same history in the tree construction. The sequence of ϵ's describes the location of the intervals in the tree structure, and assume the binary values of 0 or 1. For the circle map, the sequence $(\epsilon_{i+1}, \epsilon_i) = (1, 1)$ is forbidden, consistent with the Fibonacci sequence defining the periodic orbits. The argument of the function is determined by expanding time (or iterations) in base 2, i.e., $t = \Sigma \epsilon_i 2^i$, yielding $\sigma(t)$. After suitable normalization, this procedure associates a real number between 0 and 1 to each interval. Such an expansion also correctly organizes the intervals to be compared. It should be stressed that if one can extract the scaling function from the experiment, one has a complete microscopic description of how the attractor was refined, and one can easily compute all statistical averages.

We shall now assume that the two entities, namely the flow and the map, possess the same multiplicative structure, leading to a particular kind of scaling at finer and finer scales. We now like to quantify the refinement process by extracting the values of scale contractions at each stage of refinement and comparing them with those from the circle map. Here one runs into the difficulty that the scaling function is experimentally not a robust quantity and therefore difficult to obtain. We therefore extract a suitably modified scaling function [17], which is based on comparing intervals within a single periodic orbit rather than two orbits. Focusing on the golden mean winding number, we note that the ratios p/q, which approximate the golden mean successively better for each increasing index on the Fibonacci sequence, correspond to the different frequency-locked Arnol'd tongues mentioned earlier. We thus need to study

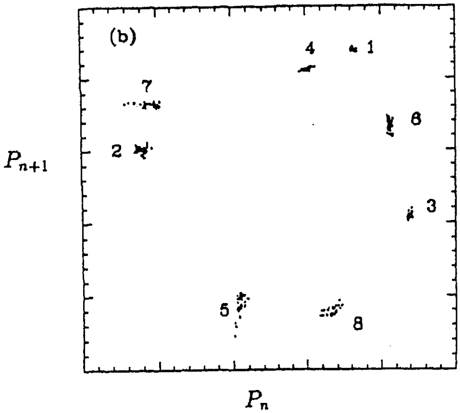


Figure 5. A typical Poincaré section for the mode-locked state for the ratio $p/q = 5/8$. The numbers by the point clusters represent the dynamical sequence of the periodic iterates creating the attractor. The sequence 1-6-3-8-5-2-7-4 is consistent with the circle map theory. The Poincaré section is embedded in three-dimensional phase space whose two-dimensional projection is presented. Three dimensional vector distances between neighboring points are used to determine the trajectory scaling function.

periodic orbits of length $F_i = q$, where $\theta_n = \theta_{n+q}$. It follows that θ_n and $\theta_{n+F_{i-1}}$ are nearest neighbors on the Poincaré section, θ_n and $\theta_{n+F_{i-2}}$ are the next nearest neighbors, and so forth. Following Ref. [17], we define the trajectory scaling function within a single periodic orbit for the circle map as

$$\sigma_j = \frac{[\theta_j - \theta_{j+F_{n-2}}]}{[\theta_j - \theta_{j+F_{n-1}}]} \quad \text{if } 0 < j \leq F_{n-2} \tag{5}$$

and

$$\sigma_j = \frac{[\theta_j - \theta_{j+F_{n-2}}]}{[\theta_j - \theta_{j-F_{n-1}}]} \quad \text{if } F_{n-2} < j \leq F_n. \tag{6}$$

Since we wish to approach the golden mean critical point with a sequence of p/q rational approximations, we experimentally adjust the system at the critical amplitude to the best possible rational approximation to the golden mean winding number. For any given p/q approximation, the Poincaré section consists of a set of q discrete points. For the experimental situation, a two-dimensional projection for $q = 8$ is shown in figure 5. The experimental trajectory scaling function is obtained using the three-dimensional vector distances between the appropriate points. The noise within each point cluster is reduced by averaging. The experimental trajectory scaling function for a 5/8 lock-in state is shown in figure 6. The reader should examine figure 6 with figure 4 of Ref. [17] to realize that the present comparison is actually of better quality than for the convection experiment considered there. The agreement between theory and experiment is reasonably good, though more data points and smaller experimental error would have been desirable.

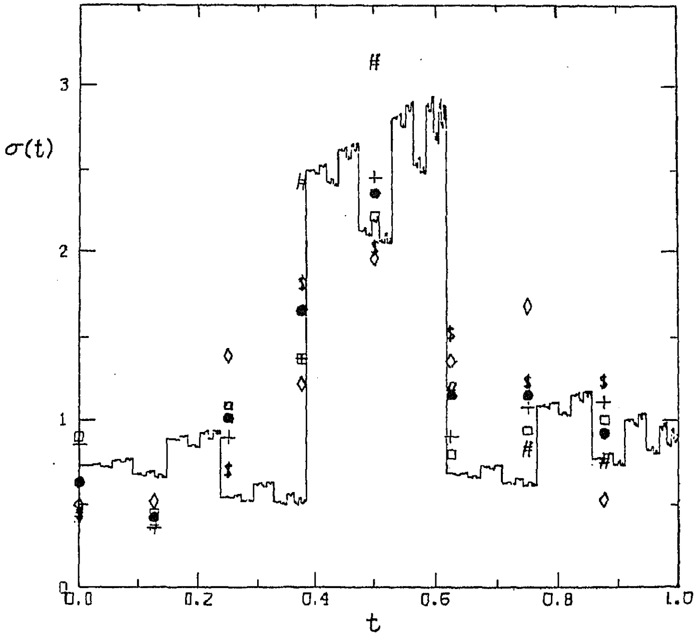


Figure 6. The trajectory scaling function from the experiment for a 5/8 lock-in state. The solid line is computed from the sine circle map at the golden mean critical point. The various symbols represent calculations from different experimental runs.

7. Conclusions

Even though some departures from the universality associated with the circle map are apparent in wake dynamics, we think that the extent of the observed similarity is remarkable. It is not obvious whether the departures observed are real or occur because the control parameters are not as fine tuned as desired. It is known, for example, that departures from criticality can produce comparable departures from universality to those observed here [18]. Apropos of this state of affairs, we reiterate that enormous care was exercised in the experiments, and state our belief that the residual problems of fine control cannot be eliminated without resorting to unconventional ways of generating such flows. A particular focus of this work has been the search for universal behavior through the study of the trajectory scaling function. This gives a detailed description of the attractor. We believe that the agreement with universality predictions for this experimentally fragile but rich function is indeed considerable.

In the past, observations such as these have evoked the question: why does a fluid flow governed by coupled field equations follow the dynamics of a simple map? The key idea here is universality: *a system may exhibit quantitative universality as long as it possesses certain common qualitative properties.* The relevant statement here is that the dynamics of many systems with two nonlinearly coupled modes, one of which is periodically driven, yield the dynamics of a sine circle map near the critical point. It must be said that at present there are no general rules that help us forecast with certainty that a flow in which a few modes are excited will or will not follow a certain universal scenario. However, with the hindsight of the experimental (or other) knowledge about the existence of a universal scenario, it is often possible to derive the simple map from the governing equations.

In the particular case of the oscillating cylinder, the experimental demonstration of the correspondence between the flow and the circle map implies that one should be able to extract the latter appropriately from the equations of motion governing the flow. Without this step, the connections implied are unsatisfying and, to a skeptical mind, perhaps no more than a curiosity. Some work in the needed direction has already occurred [11]. The starting point is the Landau-Stuart equation [19], [20] which, on the one hand, can be deduced from the equations of motion and, on the other hand, is also known to describe the wake evolution near the critical Reynolds number [21], [22]. Add to this equation an external forcing term $F = F_o \exp(i\omega_e t)$, in the same spirit as that of the oscillating cylinder experiment. From this one can generate a coupled dynamical system for the amplitude and phase of a disturbance in the forced wake. An analytical study of even this simplified dynamical system is too complex, but its numerical study shows features that are close to those observed. In the limit of low forcing, this dynamical system indeed reduces to the circle map.

Acknowledgements

The research was supported by grants to Yale University awarded by the Air Force Office of Scientific Research and the Defense Advanced Research Projects Agency. KRS appreciates the hospitality of the School of Mathematics, Institute for Advanced Study, Princeton, while completing this paper.

References

[1] H Bai-Lin, *Chaos II* (World Scientific, Singapore, 1990)

[2] M J Feigenbaum, L P Kadanoff and S J Shenker, *Physica* **D5**, 370 (182)

[3] S J Shenker, *Physica* **D5**, 405 (1982)

[4] S Ostlund, D Rand, J Sethna and E Siggia, *Physica* **D8**, 303 (1983)

[5] D J Olinger and K R Sreenivasan, *Phys. Rev. Lett.* **60**, 797 (1988)

[6] D J Olinger and K R Sreenivasan, "Universal dynamics in the wake of an oscillating cylinder at low Reynolds number," in *ASME symposium on flow induced vibrations: Nonlinear interaction effects and chaotic motions*, ASME Winter Annual Meeting, Chicago, Illinois, pp. 1–27 (1988)

[7] J Stavans, F Heslot and A Libchaber, *Phys. Rev. Lett.* **55**, 596 (1985)

[8] J Stavans, *Phys. Rev.* **A35**, 4314 (1987)

[9] M J Feigenbaum, M H Jensen and I Procaccia, *Phys. Rev. Lett.* **57**, 1503 (1986)

[10] D J Olinger, *Phys. Fluids* **5**, 1947 (1993)

[11] D J Olinger, "*Universality in the transition to chaos in open fluid flows*", Ph.D. Thesis, Yale University, 1990

[12] A Roshko, N.A.C.A. Tech. Rep. 1191, 1954

[13] See, for example, T C Halsey, M H Jensen, L P Kadanoff, I Procaccia and B I Shraiman, *Phys. Rev.* **A33**, 1141 (1986)
Aspects of multifractals were introduced in the context of strange attractors by H G E Hentschel and I Procaccia, *Physica* **D8**, 435 (1983)

In the context of turbulence, they were introduced differently by G Parisi and U Frisch, in *Turbulence and predictability in geophysical fluid dynamics*, edited by M Ghil *et al*, North Holland, pp. 719–736;

Their rudiments can be found in an even earlier paper by B B Mandelbrot, *J. Fluid Mech.* **62**, 331, 1974

For a discussion of multifractals in experimental contexts, see C Meneveau and K R Sreenivasan, *J. Fluid Mech.* 429–484 (1991)

[14] A B Chhabra, R V Jensen and K R Sreenivasan, *Phys. Rev.* **A40**, 4593 (1989)

[15] M J Feigenbaum, *J. Stat. Phys.* **19**, 25 (1978)

[16] M J Feigenbaum, *J. Stat. Phys.* **21**, 669 (1979)

[17] A L Belmonte, M J Vinson, J A Glazier, G H Gunaratne and B G Kenney, *Phys. Rev. Lett.* **61**, 539 (1988)

[18] J A Glazier, G Gunaratne and A Libchaber, *Phys. Rev.* **A37**, 523–530 (1988)

[19] L D Landau and E M Lifshitz, *Fluid Mechanics: Course in Theoretical Physics* (Pergamon Press, Oxford, 1959) vol. 6 ch. 3

[20] J T Stuart, *J. Fluid Mech.* **9**, 353 (1960)

[21] C Mathis, M Provansal and L Boyer, *J. Fluid Mech.* **182**, 1 (1987)

[22] K R Sreenivasan, P J Strykowski and D J Olinger, "Hopf bifurcation, Landau equation, and vortex shedding behind circular cylinders", in *ASME Forum on Unsteady Flow Separation,* FED **52**, 1 (1987)

PRAMANA
—journal of
physics

© Printed in India

Vol. 48, No. 2,
February 1997
pp. 705–718

Chaos in jerky flow — Experimental verification of a theoretical prediction

S J NORONHA*, G ANANTHAKRISHNA*, L QUAOUIRE+ and
C FRESSENGEAS+
*Materials Research Centre, Indian Institute of Science, Bangalore 560 012, India
+Laboratoire de Physique et Mécanique des Matériaux, U R A 1215 C N R S, Institut Supérieur de Génie Mécanique et Productique, Université de Metz, Ile du Saulcy, 57045 - Metz Cedex 01, France

Abstract. Sometime ago Ananthakrishna and coworkers had predicted the existence of chaos in jerky flow based on a nonlinear dynamical model consisting of the time evolution equations for three types of dislocations and an equation for the evolution of the stress. Our main focus here is to report the verification of this prediction by analysing the stress signals obtained from samples of AlCu alloys subjected to a constant strain rate test. The analysis of the stress signals is carried out by using several methods. The analysis shows the existence of a finite correlation dimension and a positive Lyapunov exponent. We also carry out a surrogate analysis of the time series to ascertain that the signals are not from a power law stochastic process. From the analysis we find that the minimum number of variables required for a dynamical description of the jerky flow appears to be four or five, consistent with the number of degrees of freedom envisaged in the model.

Keywords. Jerky flow; Portevin-Le Chatelier effect; dislocations; chaos; correlation dimension; Lyapunov exponent; singular value decomposition; surrogate analysis.

PACS Nos 05.45; 62.20; 81.40

1. Introduction

The work reported here is an exercise in the application of concepts and techniques in dynamical systems to problems in plasticity. The subject matter falls in the general area of 'stick-slip' phenomenon. For the sake of a general audience, we very briefly recall some basic features of a specific form of plastic instability arising under a constant strain rate test. This is known as jerky flow or the Portevin-Le Chatelier (PLC) effect and has been an object of attention for a long time in metallurgical literature. Dislocations are the basic defects contributing to the plastic deformation of specimens. The stress–strain curves show repeated stress drops and each stress drop in the deformation curve is associated with the occurrence of a dislocation band. The band usually propagates along the sample. The phenomenon is seen in a number of alloys and is known to manifest only in a window of applied strain rates and temperatures [1]. Alloys always contain point defects such as substitutional and interstitial atoms which are commonly referred to as solute atoms. At intermediate temperatures, the cores of these dislocations are saturated with solute atoms, since these atoms have a tendency to migrate to the cores of the dislocations. Consider a constant strain rate experiment and assume that dislocations are initially free from solute atoms. Once the stress developed in the sample is enough to move dislocations, they sweep the sample and in the process gather solute atoms leading to immobilization of dislocations. This happens once the solute concentration at the core of the dislocation exceeds a certain critical value. However, since the sample is subject to a constant strain rate deformation, the stress in the sample increases to such a value that dislocations break away from the clouds of solute atoms. The breakaway of dislocations results in an yield drop. This whole process repeats thereby multiple yield drops are seen. It is clear that this repetitive nature of the process arises due to a competition of two time scales, namely, the time scale associated with the velocity of dislocations and the diffusion time scale of the solute atoms. This phenomenon is known as dynamic strain ageing (DSA) and was first envisaged by Cottrell [2]. In spite of the enormous attention the subject has received, several aspects of the phenomenon were not well understood until recently. There has been renewed attempts [3–9] to understand this phenomenon by introducing concepts borrowed from the theory of dynamical systems. This has helped to gain new insights hitherto not possible. In most of the models, the negative strain rate sensitivity (SRS) is an input into diffusion like equations [8,9]. Some of these models have been quite successful in explaining a number of features of the PLC effect including the band velocity. Even the existence of chaotic solutions remains a possibility [9]. The first truly dynamical model was attempted by Ananthakrishna and coworkers [3–5] several years ago, wherein the spatial aspect were ignored. This method starts from the time evolution equations for the densities of three types of dislocations. These equations are then coupled to the equation for the stress developed in the sample. The negative SRS results as a consequence of a Hopf bifurcation from a time homogeneous steady state to the time oscillatory state [3–5,12]. The theory has been improved to include the spatial dependence as well [13]. Although, the original theory ignored the spatial aspects, it proved to be surprisingly successful in that it could explain several features of the PLC effect [5,12].

In addition to above results, *one specific prediction of the model is the possibility of chaotic behaviour in a limited range of strain rates* [10, 11]. The first detailed report of the analysis of the experimental stress-strain series which demonstrated the existence of chaos was published recently [14], eventhough, this result was quoted in an earlier paper [13]. Since then there has been a few other investigations which have provided additional support that jerky flow is chaotic [15–17]. *The implication is that the dynamical basis of the model with only a few degrees of freedom is correct* [14].

The purpose of the paper is to present additional evidence that the dynamics of the underlying phenomenon is indeed chaotic. The methods include the singular value decomposition, calculation of a finite correlation dimension and a positive Lyapunov exponent. In addition, we carry out surrogate analysis to demonstrate that the time series is not due to a power law stochastic process. The plan of the paper is as follows. In §2, we briefly present the model for the PLC effect which predicts the existence of chaos along with the results related to chaos. In §3, we outline the various methods of analysis used here. We then apply them to the experimental time series. Section 4 contains discussion and conclusions.

2. Chaos in a model for the PLC effect

In the following, we very briefly outline our model for the PLC effect [3–5] in terms of scaled variables. Details of the model can be found in the original papers. It is a dynamical model using three types of dislocation densities, namely the mobile dislocation density denoted by x, the immobile dislocation density denoted by y, and another type which may be regarded as dislocations with clouds of solute atoms denoted by z. The basic mechanisms included in the model are: dislocation multiplication through cross glide, annihilation of dislocations, formation of dislocation locks, gradual immobilization of mobile dislocations by solute atoms, and the re-mobilization of the immobile dislocations by stress or thermal activation. These mechanisms lead to a coupled set of nonlinear differential equations for the rate of change of the densities. These rate equations are coupled to the machine equation. Using a power law dependence for the velocity of the mobile dislocations in terms of scaled stress ϕ given by $V_m = V_0(\phi)^m$, we can write the equations for the time evolution of dislocation densities and the scaled stress ϕ as

$$\dot{x} = \phi^m x - b_0 x^2 - ax + y - xy, \tag{1}$$

$$\dot{y} = b_0[\kappa b_0 x^2 - xy - y + az], \tag{2}$$

$$\dot{z} = c[x - z], \tag{3}$$

$$\dot{\phi} = d[e - \phi^m x], \tag{4}$$

where the parameter a refers to the concentration of the solute atoms, b_0 to the strength of reactivation of the immobile dislocations, c to the time scales over which the slowing down of dislocation occurs, d to the effective modulus and e to the applied strain rate. There is a range of values of the parameters a, b, c, d, κ, e and m for which we find oscillatory solutions. The model is known to produce a large number of qualitative features of the PLC effect [3–5]. One specific prediction which is of interest here is that the model exhibits chaos in an intermediate range of applied strain rates [10–12]. The model shows the period doubling route to chaos when the velocity exponent $m \geqslant 2$ [11,12]. The value of the Feigenbaum's exponent is the same as that for the quadratic map. For $m > 2$, it shows a bubble structure. For m close to unity, the model shows a period adding phenomenon. Other studies carried out on the model include obtaining approximate closed form expressions for the limit cycles [4] and deriving the time dependent Ginzburg-Landau equation for the order parameter for the creep case [18] (described by equations 1–3). The model has since been extended to include spatial coupling [13]. An additional result that emerges is the velocity of the band.

3. Detection of chaotic behaviour in experiments

3.1 *Experimental procedure*

Experiments performed under constant strain rate conditions often exhibit two features that are not desirable from the point of dynamical analysis. First, there is an upward drift of the stress-strain curve. The physical origin of this lies in the fact that more and more dislocations get immobilized and are stored in the sample with a time scale commensurate with the time scale of immobilization. This is called as strain hardening. In addition, there is a gradual increase in the amplitude of the serrations. Clearly both these features are undesirable since they imply nonstationarity of the time series. However, it is often possible to find stress–strain curves which are flat. In the present case, the stress signals are in such a region. The experiments were performed on single crystals of CuAl 14% oriented along $< 111 >$ at 290 K at Poitiers. The strain rate of deformation was $3.34 \times 10^{-4}\mathrm{s}^{-1}$. The stress signals were recorded in a digitized form from a single deformation curve. Each set contained 4720 points recorded at a rate of 242 points per second. Even though the data sets are short, we have twelve data sets. As an illustrative example, we will present the results on one data set.

3.2 *Time series analysis*

It is known that the stress drops in constant strain rate experiments can appear noisy in a certain regime of the parameter space and devoid of any pattern. Whenever one meets such a noisy sequence, the main objective of any analysis is to determine whether it is of deterministic origin or stochastic origin. In general stochastic noise is always present, for instance, the machine noise or inherent randomness due to other causes.

Our first analysis of the experimental data on Al–Cu alloys [14] has shown that there is a finite correlation dimension. This was obtained using the algorithm due to Grassberger and Procaccia (GP) [19]. However, the existence of a finite correlation dimension cannot always be taken as an indicator of chaotic dynamics. This is because a stochastic process with a power law spectrum can also exhibit a finite correlation dimension [20, 21], and therefore such a process mimics a low dimensional attractor. There are other methods such as singular value decomposition [22] which can be used effectively, particularly when used in conjunction with other methods [23]. Even so, these methods only provide dimensional estimates, and as such cannot always be taken as an evidence of chaotic behaviour. The most reliable quantifier of chaotic dynamics is the existence of a positive Lyapunov exponent. However, conventional methods of determining the Lyapunov exponent require long time series [24, 25] which are impractical to obtain from experiments. Recently, the well known Eckmann's algorithm has been modified to suit short time series [26] giving the entire Lyapunov spectrum. Yet another particularly simple method is due to Gao and Zheng [27]. This method allows one to calculate both the optimum delay time and the optimum embedding dimension in addition to providing an estimate of the positive Lyapunov exponent. Finally, it is always worthwhile to generate surrogate data sets and repeat the analysis to ensure that the signals are not due to a power law stochastic process.

All the methods are based on phase space reconstruction by embedding the signal in a higher dimensional space and carrying out the required analysis. The purpose of embedding a scalar time series in higher dimension is to reconstruct other variables which constitute the dynamics of system from the measured time series. This is done by embedding the signal in higher and higher dimension until an optimum dimension is reached. The object of increasing the embedding dimension is to remove or minimize the contribution from 'false neighbours' since at small embedding dimensions, there are large number of such 'false neighbours'. The rule of thumb for selecting the time lag is to use

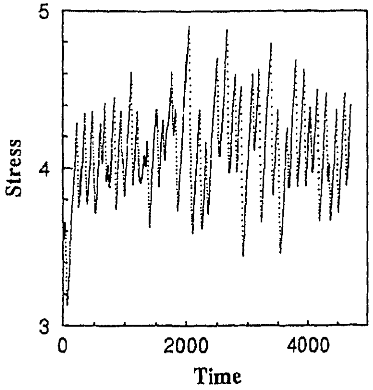


Figure 1. A typical stress-time plot for AlCu alloy.

the value of the time at which the correlation function falls to $1/e$ of the initial value. Some use a value of the first zero crossing time. Thus, there is some arbitrariness in estimating the optimal time delay. The method due to Gao and Zheng [27] allows one to determine the optimum value of the time lag as well as the embedding dimension (see below).

Consider a stress signals measured from a jerky flow normally sampled at regular intervals of time represented by $\{\sigma_i = 1, \ldots, N\}$. A typical stress-strain curve is shown in figure 1. The constructed d dimensional vector is $\vec{\xi_i} = \{\sigma_i, \sigma_{(i+\tau)}, \ldots, \sigma_{[i+(d-1)\tau]}\}$, where τ is some fixed time interval used as a delay. Noise present in the signal can be 'cured' by using singular value decomposition technique. In this method, the (d×N) trajectory matrix A

$$A^t = \{\vec{\xi}_1, \vec{\xi}_2, \ldots, \vec{\xi}_N\} \tag{5}$$

is rotated onto the basis of its principal vectors. The d principal values of the matrix A (ie., the eigenvalues of the covariance matrix A^tA) are positive. Eigenvalues are conventionally ordered in a decreasing order. If some of the principal values say, $(q+1, \ldots, d)$ are zero, then, it is clear that the trajectory remains confined to the subspace spanned by the basis $(1, \ldots, q)$. In practice the presence of noise prevents these eigenvalues from taking zero value exactly. However, there may be a set of eigenvalues which are small compared to the largest one. Then, an estimate of the embedding dimension is obtained by the sharp decrease in the eigenvalue to a certain low level. It is possible then to obtain the cured trajectory matrix by retaining only the first q components and back rotating these components. Often, this method is used as a noise reduction technique.

The correlation integral $C(r)$ is calculated by

$$C(r) = \frac{1}{N_p} \sum_{i,j} H(r - |\vec{\xi}_i - \vec{\xi}_j|) \tag{6}$$

where $H(\cdot\cdot)$ is the Heaviside step function and N_p is the number of pairs of the embedding vectors $(\vec{\xi}_i, \vec{\xi}_j)$ whose distance is less than r. These vectors may refer to either the d dimensional vectors obtained directly from the time series or to the d dimensional vectors obtained by back rotating the first p principal components. We still need to identify the optimal delay time. This can be done using the following method.

In a chaotic time series, points on two neighbouring trajectories diverge exponentially in time. This divergence can be measured by calculating the distance between them at $t = 0$ and at a later time after k time steps. In a short time the difference vector corresponding to the two points aligns itself in the direction of maximum stretching. Consider two vectors $\vec{\xi}_i$ and $\vec{\xi}_j$ in d dimension. Let $d_{i,j}(k) = \| \vec{\xi}_{i+k} - \vec{\xi}_{j+k} \|$, where $\| \cdot\cdot \|$ is the Euclidean distance between the two vectors at time k between $\vec{\xi}_i$ and $\vec{\xi}_j$. Let $d_{i,j}(0)$ be the initial distance. Since for a chaotic system, $d_{i,j}(k)$ should be larger than $d_{i,j}(0)$, a measure of the divergence is the time dependent Lyapunov exponent $\Lambda(k, d, \tau) = \langle \ln[d_{i,j}(k, z)/d_{i,j}(0)] \rangle$, where $\langle \cdot\cdot \rangle$ refers to the average over all pairs of points for which $d_{i,j}(0)$ is less than a chosen d_0. Another quantity is $\Lambda_+(k, d, \tau)$ where the average is limited to those pairs of points for which $(d_{i,j}(k)/d_{i,j}(0)) > 1$. An appropriate window $t_w = (d-1)\tau$ is imposed such that points on the same orbit satisfying $|i - j| > t_w$ are discarded. The optimal time lag is found by looking for the minima in the plots of $\Lambda(\tau, d, k)$ vs τ and $\Lambda_+(\tau, d, k)$ vs τ for various values of the embedding dimension. The optimum embedding dimension can also be determined by looking for minimal changes in $\Lambda(\tau, k, d)$ versus τ as d is increased around the minimum of the curve. In addition, the divergence plots of $\ln[d_{i,j}(k)/d_{i,j}(0)]$ vs $\ln[d_{i,j}(0)]$ do not change much as we increase the embedding dimension beyond a certain dimension which is identified as the optimal embedding dimension.

Since the time series is short, the calculation of the entire Lyapunov spectrum is limited to a few embedding dimensions ($d < 5$). This method is a modification of the algorithm due to Eckmann *et al* [25]. It relies on the construction of a sequence of (d_M, d_M) matrices T_i, $(d_M \leqslant d)$ which map the difference vector $\vec{\xi}_i - \vec{\xi}_j$ to $\vec{\xi}_{i+k} - \vec{\xi}_{j+k}$ and successively reorthogonalize T_i using the standard $Q_i R_i$ decomposition. (Q_i is an orthogonal matrix and R_i is an upper triangle matrix with positive diagonal elements.) Then the Lyapunov exponent is given by

$$\lambda_l = \frac{1}{kp\Delta t} \sum_{j=0}^{p-1} \ln(R_j)_{ll}, \quad l = 1, 2, \ldots, d \tag{7}$$

where p is the number of available matrices and k is the propagation time in units of Δt which is the time step. The modification effected in this method is to choose all orbits in a shell rather than a sphere so that the effect of randomness is eliminated. The time lag used in this algorithm is the value where the correlation function falls to $1/e$ of the initial value or an optimum choice determined by another method.

Finally, one method of determining if the signals are due to a stochastic process with a power law spectrum is to construct surrogate data sets and repeat the analysis. Surrogate data sets can be generated in a number of ways [21]. One simple way is to take the Fourier transform of the signals randomize their phases and Fourier invert it. If the surrogate data sets do not show a finite correlation dimension or a positive Lyapunov exponent, while the original data sets exhibit both, this is taken as a confirmation that the data is due to a low dimensional attractor.

All the algorithms have been checked against standard benchmark cases such as Rossler and Lorenz attractor; good agreement with the known values in the literature has been obtained.

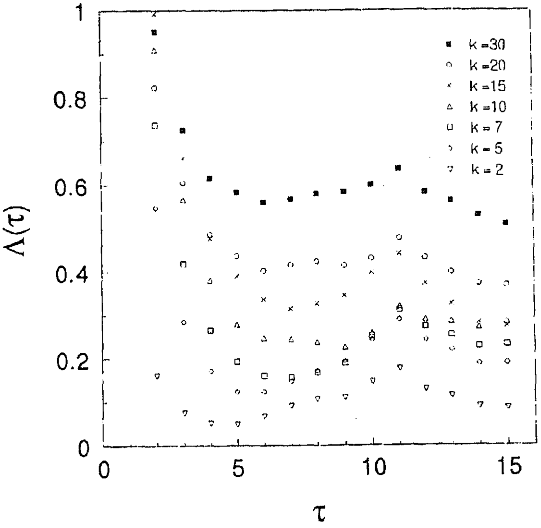


Figure 2. A plot of $\Lambda(k, \tau)$ vs τ for various values of k for the Rossler attractor.

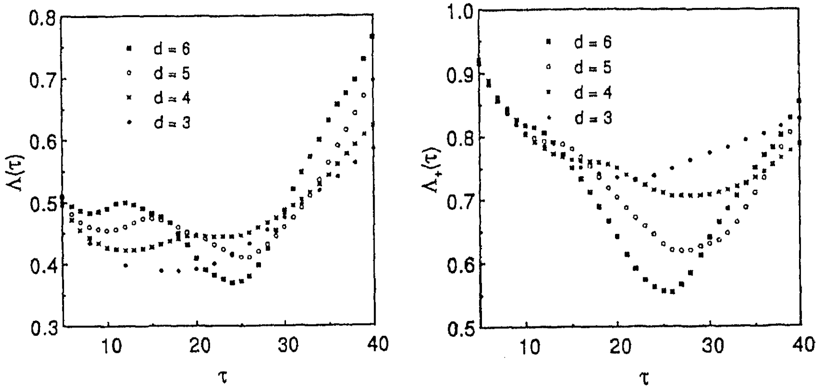


Figure 3a,b. A plot of $\Lambda(\tau)$ and $\Lambda_{+}(\tau)$ vs τ for $k = 20$ for various values of d.

3.3 *Analysis of experimental signals*

In the method due to Gao and Zheng, the choice of an appropriate value of k is not specified. Generally, it is so chosen that it reflects a reasonable extent of divergence and is taken to be of the order of the correlation time. However, often one finds that the minima in $\Lambda(\tau, k, d)$ vs τ coincides with the value of k chosen and therefore gives a spurious optimal lag time. We found that it is possible to estimate the correct value k by plotting $\Lambda(\tau)$ vs τ for various values of k. We find that the minima in $\Lambda(\tau)$ vs τ for various values of k initially increases as k is increased and then decreases going through a maximum. This maximum value is chosen to be the optimum value of k. A plot of $\Lambda(\tau, d, k)$ vs τ for the Rössler attractor is shown in figure 2 for various values of k. The optimum k can be seen to be 9. We find that this value k_{opt} is independent of the choice of d_0.

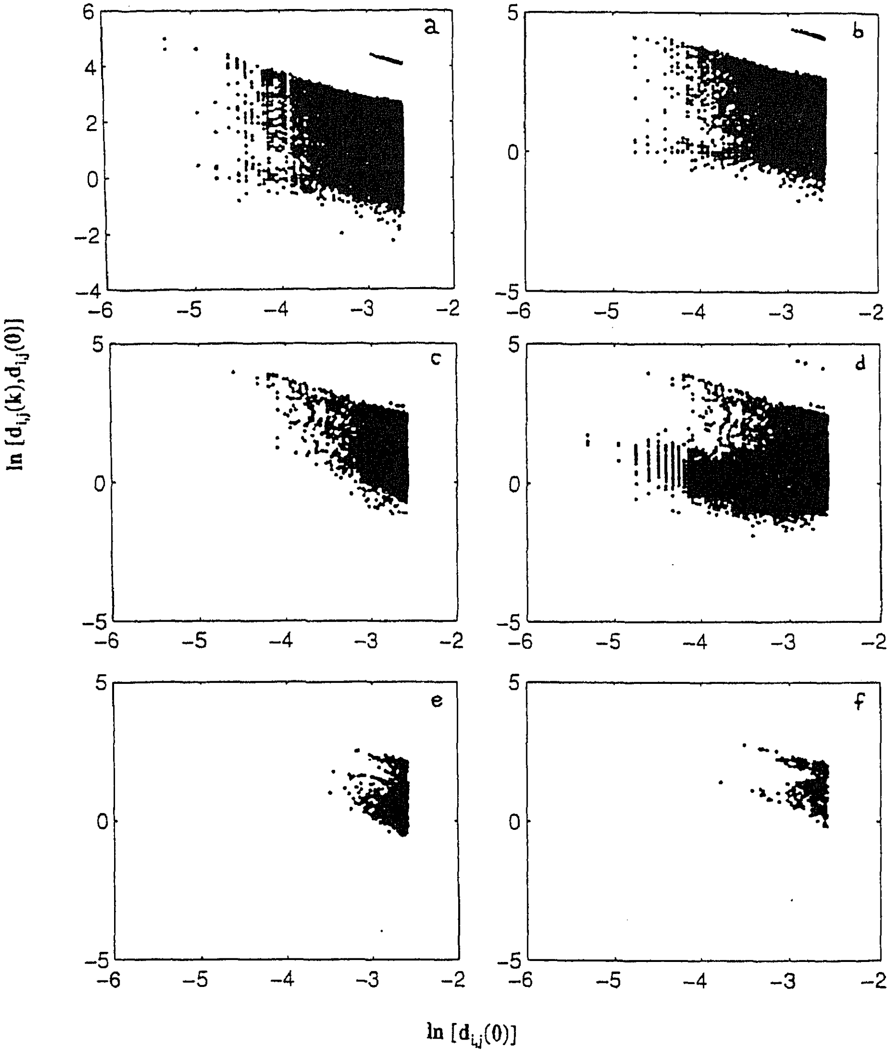


Figure 4. The local divergence plots with a window: **(a)** $d = 4, \tau = 27$, **(b)** $d = 5, \tau = 27$, **(c)** $d = 5, \tau = 27$, **(d)** $d = 5, \tau = 27$ (without a window), **(e)** $d = 6, \tau = 27$, **(f)** $d = 6, \tau = 37$.

Figures 3a–b show plots of $\Lambda(\tau, d)$ and $\Lambda_+(\tau, d)$ respectively for $d = 4, 5$ and 6 for $k = 20$ for one data set. When we increase the dimension form 4 to 6, we see that the curve corresponding to $d = 5$ lies lower than $d = 4$, while $d = 6$ does not change significantly from $d = 5$. This suggests that $d = 5$ is the optimum dimension. Further, we also note that the minimum occurs around $\tau = 27$, corresponding to the optimal time lag. Figures 4a–f show the divergence plots for various embedding dimensions and delay times. It is clear that as we approach the optimal d, the divergence plots become most compact. Thus, the optimal embedding dimension appears to be $d = 5$. Nonoptimal

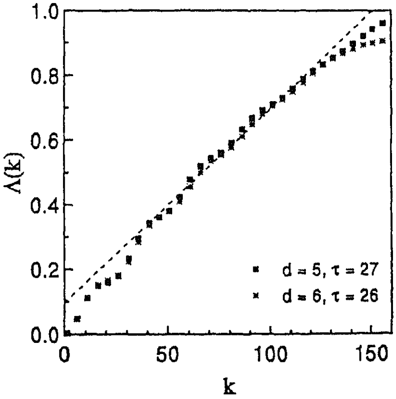


Figure 5. A plot of $\Lambda(k)$ vs k corresponding to two different sets of d and τ.

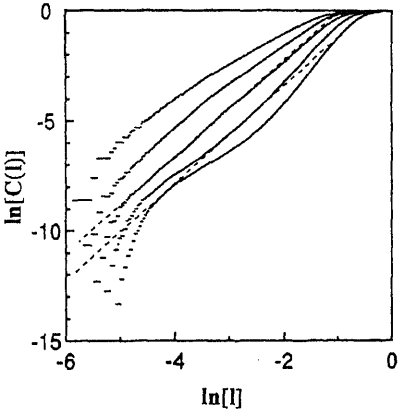


Figure 6. Plot of the correlation integral for different d for $\tau_{opt} = 27$.

values of the time lag τ can be seen to give rise to excess positive excursions over the optimal value. This can be seen from figure 4f shown for $d = 6$ and $\tau = 37$. Note that most of the values are positive even though the plot appears quite compact. Using the value of the optimal time delay, we calculate the time dependent Lyapunov exponent $\Lambda(\tau, k, d)$. This is shown in figure 5 for $d = 5$ and 6. We note that the curve corresponding to $d = 5$ is slightly more linear and has a small positive intercept. The value of the Lyapunov exponent obtained from this plot is 1.4. Having obtained the optimal time delay, we can use this to compute the correlation dimension for this data using the G-P algorithm. This is shown in figure 6. It is clear that even for very small embedding dimension of 5 there is a saturation effect of the slope. Note the dashed straight lines drawn for $d = 4$ and 5 are parallel implying the convergence in the slope. The slope thus obtained is 2.1. It may be worthwhile to point out that the shrinking of the scaling regime starts at fairly small values of r for $d = 6$. Beyond this embedding dimension we find that the scaling regime becomes very small. This behaviour is common to short time series. The saturation of the slope for such a low embedding

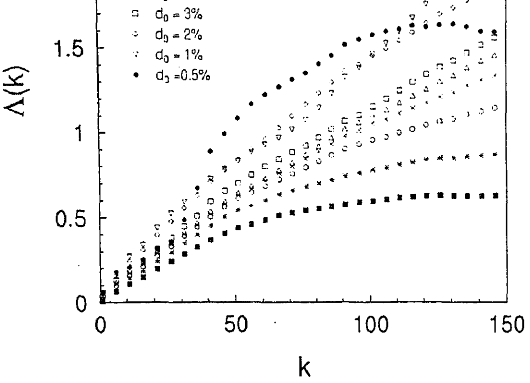


Figure 7. A plot of $\Lambda(\tau, k, d)$ vs k for optimum τ for various values of d_0 for the Rossler attractor.

dimension must be contrasted with the high embedding dimension required for a similar saturation in our earlier calculation where a much shorter time lag was employed. This can be clearly attributed to the small delay time that was used in our previous calculation [14] since it is the total window length t_w that remains constant.

It is worthwhile to comment here on the actual values of the Lyapunov exponent obtained by this method. We found that the Lyapunov exponent depends on the value of d_0 chosen. Using large values of d_0 gives smaller values of the exponent, since the number of points contributing to the average is large, and often, is not a true indicator of the divergence. On the other hand, if too few points are included (ie., d_0 is small) then the statistics gets poor as is the case for $d_0 = 0.05\%$ of the attractor. Indeed, the extent of linearity appears to be less when d_0 is large and it improves as d_0 is reduced. It appears to break down for small values. Thus a choice of a few percent appears to be reasonable. This is shown in figure 7 for the case of the Rossler attractor. The value of the Lyapunov exponent obtained by this method is a lower bound to that obtained via calculations (which give the largest exponent [24] or the entire spectrum), since in the present method the average is over all possible values including the contracting directions. Even so, we feel that fixing an appropriate value of this parameter in a way that gives a correct value of the Lyapunov exponent for the experimental signals is not easy due to the presence of noise and therefore, the actual value cannot be taken seriously. However, it must be emphasized that the Lyapunov exponent is positive. This is an adequate criterion for the time series to be chaotic.

The singular value decomposition technique has been used to analyse the data as well. The spectrum for various values of window length t_w is shown in figure 8 . It is clear that there is a sharp drop in the relative magnitude of the eigenvalues. Indeed, the maximum fall is at the second component. The changes there after are not so large. This shows that the number of degrees of freedom is of the order of two or three, consistent with the value

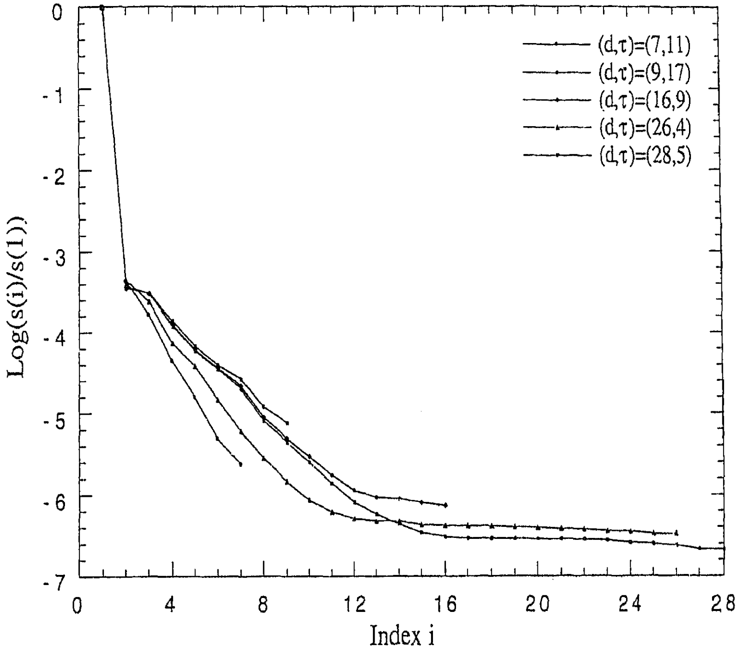


Figure 8. A plot of the singular value spectrum for various values of window length.

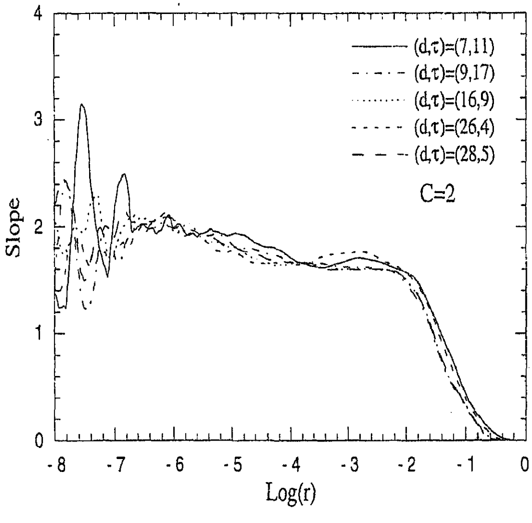


Figure 9. A plot of slope of the correlation integral for various values of window length retaining only the first two principal components.

obtained from calculation of the correlation dimension. As an illustration of the method, we can estimate the correlation dimension on the cured time series obtained by back rotating by keeping only the first few components. Figure 9 shows the result obtained by keeping only two components. The convergence of the slope is clearly around 2. General

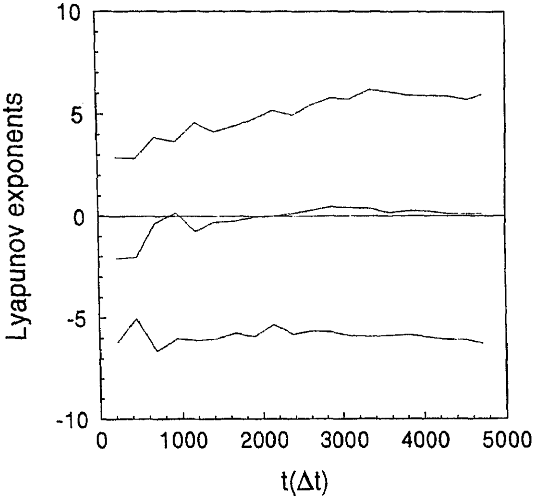


Figure 10. A plot of the Lyapunov spectrum as a function of time for $d = 3$ for the original data.

features are not altered if more components are included, since the correlation dimension when all the components are considered is 2.1 as seen from figure 6.

The spectrum of Lyapunov exponents have been calculated using a delay time of $\tau = 27$ for the first few embedding dimensions. From $d = 3$, we find a positive exponent. A plot of the Lyapunov spectrum as a function of time (in units of Δt) is shown for $d = 3$ in figure 10. The value of the largest, the smallest and the zero exponent remains roughly the same from $d = 3$ up to $d = 5$. However, there is an additional positive Lyapunov exponent seen for $d = 5$ which is spurious. This statement is based on calculations carried out on standard models like the Lorenz and the Rossler attractor. In these two cases, we find that at $d = 5$, there is an additional spurious positive eigenvalue. This is due to the fact that we need at least 10000 points for $d = 5$.

Finally, we have generated six surrogate data files. We have calculated both the correlation dimension and the Lyapunov spectrum. We find no saturation in the slope of the scaling regime in any of the six surrogate data sets. Further, we do not see any positive Lyapunov exponent for dimensions up to $d = 5$ for any of the six surrogate files. This clearly shows that the underlying dynamics of the original data is due to a low dimensional attractor.

4. Discussion and conclusions

Using a combination of several methods we have analysed time series obtained from experiments on AlCu alloys. The singular value decomposition method clearly shows that the orbit is confined mostly to a few dimensions. We have used a method due to Gao and Zheng to obtain the optimum time delay and the optimum embedding dimension. Since the method gives positive values for the Lyapunov exponent, one can confidently say that the jerky flow is chaotic. In addition, using the optimum value of the time lag, we have calculated the correlation dimension of value 2.1. We also have calculated the spectrum of Lyapunov exponents by a method suited for short time series. There is one positive Lyapunov exponent of magnitude ~ 4. We have further carried out a surrogate analysis to show that the time series is due to a low dimensional chaotic attractor. These calculations confirm the theoretical prediction that jerky flow could be chaotic. Further, the analysis shows the optimum embedding dimension for the time series to be $d_{\rm opt} = 5$. This is also consistent with the correlation dimension 2.1 obtained by using optimal time delay of $\tau_{\rm opt} = 26$. *In conclusion, the important prediction, namely, the existence of chaotic flow in the PLC effect is confirmed. This verification strongly suggests the inherent correctness of the basis of the theory, namely, that the dynamical origin of the PLC effect is the nonlinear interaction between the participating defects. Further, the optimum embedding dimension being 5 tells us that we need only five degrees of freedom to describe the dynamics of jerky flow.* We recall that we had only four degrees of freedom since the variable corresponding to the concentration of solute atoms was deliberately suppressed in the original formulation of the model for the sake of simplicity. Lastly, it may be recalled that the phenomenon is spatially inhomogeneous which implies infinite degrees of freedom. Thus, these few degrees of freedom must correspond to the collective behaviour of dislocations and other defects. This fact gives a clue as to why this theory works so well in spite of the fact that it ignores the spatially inhomogeneous aspect which is thought to be so crucial for a proper description of the phenomenon.

Acknowledgement

This work is partly supported by IFCPAR Grant No. 1108-1. One of the authors (GA) wishes to acknowledge the support from IUFM de Lorraine and the Univeristé of Metz. The authors are thankful to the Poitier's group for supplying the data.

References

[1] B J Brindley and P J Worthington, *Metal. Rev.* **145**, 101 (1970)
[2] A H Cottrell, *Phil. Mag.* **44**, 829 (1953)
[3] G Ananthakrishna and D Sahoo, *J. Phys.* **D14**, 2091 (1981)
[4] M C Valsakumar and G Ananthakrishna, *J. Phys.* **D16**,1055 (1983)
[5] G Ananthakrishna and M C Valsakumar, *J. Phys.* **D15**, L171 (1982)
[6] L P Kubin and Y Estrin, *Acta Metall.* **33**, 397 (1985)
[7] L P Kubin and Y Estrin, *Acta Metall.* **38**, 697 (1990)
Y Estrin and L P Kubin, *J. Mech. Behavior of Materials* **2**, 255 (1989)
[8] Viewpoint Set on Propagative Instabilities, *Scripta Metall. Mater.* **29**, 1147 (1993)
[9] V Jeanclaude, C Fressengeas and L P Kubin, *Non Linear Phenomena in Materials Science II: Solid State Phenomena*, Vol 23–24, (Switzerland: Trans Tech Publications 1992) pp. 403.
[10] G Ananthakrishna and M C Valsakumar, *Phys. Lett.* **A95**, 69 (1983)
[11] T M John and G Ananthakrishna in *Directions in Chaos* Ed. Hao Bai-Lin, (Singapore: World Scientific 1990), pp 133
[12] G Ananthakrishna in *Non Linear Phenomena in Materials Science II: Solid State Phenomena*, Vol 23–24, (Switzerland: Trans Tech Publishers 1992) pp. 417
[13] G Ananthakrishna, *Script. Met.* **29**, 1183 (1993)
[14] G Ananthakrishna, C Fressengeas, M Grosbras, J Vergnol, C Engelke, J Plessing, H Neuhäuser, E Bouchaud, J Planés and L P Kubin, *Scripta Metall. Mater.* **32**, 1731 (1995)

[15] G Ananthakrishna and S J Noronha, *Solid State Phenomena*, Vol 42–43, (Switzerland: Scitech Publishers 1995), pp. 277
[16] L Quaouire and C Fressengeas, *Solid State Phenomena*, Vol 42–43, (Switzerland: Scitech Publishers 1995), pp. 293
[17] S Venkadesan, K P N Murthy and M C Valsakumar, *Solid State Phenomena*, Vol 42–43, (Switzerland: Scitech Publishers 1995), pp. 287
[18] M Bekele and G Ananthakrishna, in, *Solid State Phenomena*, Vol. 42–43, (Switzerland: Scitech Publications 1995) pp. 303; M Bekele and G Ananthakrishna, submitted to *Physica D*
[19] P Grassberger and I Procaccia, *Physica* **D9**, 189 (1983)
[20] A R Osborne and A Provenzale, *Physica* **D35**, 357 (1989)
[21] J Theiler, S Eubank, A Longtin, B Galdrikian and J D Former, *Physica* **D58**, 77 (1992)
[22] B S Broomhead and G P King, *Physica* **D20**, 217 (1986)
[23] A M Albano, J Muench, C Schwartz, A I Mees and P E Rapp, *Phys. Rev.* **A38**, 3017 (1988)
[24] A Wolf, J B Swift, H L Swinney and J A Vastano, *Physica* **D16**, 285 (1985)
[25] J P Eckmann, S O Kamphorst, C Ruelle, and S Cileberto, *Phys. Rev.* **E34**, 4971 (1986)
[26] X Zeng, R Eykholt and R A Pielke, *Phys. Rev. Lett.* **66**, 3229 (1991)
[27] J Gao and Z Zheng, *Phys. Rev.* **E49**, 3807 (1994)

—Journal of
physics

February 1997
pp. 719–736

Chaos and predictability of the Indian summer monsoon

B N GOSWAMI
Centre for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bangalore 560 012, India

Abstract. Predictability of the Indian summer monsoon is investigated by conducting three multi-year integrations with the Geophysical Fluid Dynamics Laboratory's climate model. The mean monsoon simulated by the model is realistic. It is shown that a significant fraction of the interannual variance of the simulated Indian summer monsoon may be due to internal dynamics. It is discovered that the tropical atmosphere is capable of sustaining a quasi-biennial oscillation (QBO) accounting for most of the internal low frequency variability. It is also shown that neither air-sea interaction nor surface hydrology feedback is essential for the QBO of the model atmosphere. That such a QBO can arise due to modulation of the nonlinear intraseasonal oscillations by the annual cycle is demonstrated using a simple nonlinear dynamical model. The phase and the amplitude of the internal mode is unpredictable and hence may be responsible for limiting the long range predictability of the monsoon.

Keywords. Seasonal monsoon; predictability; slowly varying forcing; internal dynamics.

PACS Nos 47.27; 92.10; 92.20; 92.60; 95.75

1. Introduction

The weather representing the instantaneous state of the atmosphere and the climate representing the space/time averages are both governed by a set of physical laws. The atmosphere and the coupled ocean-atmosphere system may be represented by a set of equations that include certain approximations in describing some of the physical processes such as radiative heating, convective heating associated with clouds, boundary-layer turbulent fluxes of heat, momentum and mass etc. With initial conditions defined by observations, such a set of equations or a 'model' may be integrated forward in time using a powerful computer. With improved models and improved specification of the initial conditions, what limits our ability to predict the weather? This question was first addressed by Lorenz [1]. Using a simple model, he showed that small errors in the initial conditions would double in about three days. Keeping in mind some of the intrinsic limitations of the observing system, this puts a limit of two to three weeks on our ability to predict the weather. This limit, often known as the 'limit on deterministic predictability', has now been confirmed using more sophisticated models. This limit is governed by the instabilities and nonlinearities of the system. The instabilities make the initial error grow in time while the nonlinearities help the errors to reach a saturation level keeping the system within certain bounds.

In contrast to the tremendous strides made by numerical weather prediction, the climate prediction efforts using deterministic models are still in early stages of deterministic limit on predictability could be achieved. However, it was demonstrated by Charney and Shukla [2] that although the instantaneous state of the atmosphere (or the weather) may not be predictable beyond 2–3 weeks, the climate or the statistical properties of the atmosphere (such as time and/or space means) may still be predictable beyond this limit. This is possible because the climate is governed by low frequency planetary-scale flow patterns. If these low frequency (LF) planetary-scale flow patterns were solely governed by internal dynamics or instabilities of their own scale or scale interactions with the high-frequency synoptic disturbances, then there will not be much hope for predictions beyond the limit on deterministic predictability. However, these LF planetary-scale patterns are also expected to be governed to a large extent by the forcing associated with slowly varying boundary conditions such as sea surface temperature (SST), soil moisture, snow cover, etc. These slowly varying boundary conditions give rise to slowly varying forcings in the atmosphere that provide significant coherent changes in the planetary scales at LF, leading to potential predictability of the space-time averages. This has been supported by a large number of atmospheric general circulation model (AGCM) studies forced by observed sea surface temperature (SST) variations (see [3] and references therein) and provide the conceptual basis for climate prediction.

Several observational and modeling studies have established that the Indian summer monsoon is linked with several surface boundary forcings. In addition to links with Eurasian snow cover [4–6] and soil moisture over Indian continent [7], the link with Pacific SST and Southern oscillation is considered strong and has been studied by many authors [8–11]. This background led to the optimism that long range dynamical prediction of the Indian summer monsoon one or two seasons in advance may be feasible. As a result, a large number of studies in the last few years have made serious attempts to simulate the interannual variability (IAV) of Indian summer monsoon [12–14] and to hind-cast seasonal mean monsoon rainfall [15, 16] using different atmospheric general circulation models (AGCMs). Three important observations made from these studies are summarized below.

First, most models still have a serious problem in simulating the mean summer precipitation over the Indian region. They tend to simulate too little precipitation over the continent. In some cases, even during the summer months a strong inter tropical convergence zone (ITCZ) persists over the equatorial Indian ocean. Second, most models can simulate the interannual variability of some planetary scale component of the monsoon such as the divergent circulation (as given by the velocity potential). However, the simulation of the IAV of the monsoon rainfall differs widely from one model to another [15] indicating great sensitivity of this regional part of the circulation on resolutions and physical parameterizations of the models. In addition, while the prediction of the seasonal mean rainfall in other parts of the tropics (e.g., Sahel or equatorial Pacific) do not seem to be sensitive to small changes in the initial conditions, the simulation of the seasonal mean Indian monsoon rainfall seems to be rather sensitive to small changes in the initial conditions [17]. This indicates that the mean monsoon circulation in the tropics may be unique in that it is not entirely forced by slowly varying boundary forcings but is also governed by internal dynamics to some extent.

The limit on predictability of Indian summer monsoon would be determined by the degree to which the mean monsoon is governed by internal dynamics. Quantitative estimates of this contribution to the mean monsoon is currently lacking. In the present study, an attempt is made to make quantitative estimates of contributions from boundary forcings and internal dynamics to the interannual variability of the monsoon. As our primary objective relates to the predictability of the Indian summer monsoon, we select an AGCM whose simulation of mean precipitation during the northern summer over the monsoon region is reasonably close to the observed mean precipitation. Then, the forced variability is estimated from a multi-year integration of the model with observed sea surface temperature as the boundary condition. The internal variability is estimated from a multi-year run of the same model with fixed seasonal cycle SST. Another sensitivity experiment with the fixed seasonal cycle of SST as well as that of the soil moisture has been conducted to derive some insight regarding the origin of the low frequency internal oscillations in the model. The model used and the experiments conducted are described in § 2. The mean monsoon simulation of the model is discussed in § 3. The simulation of the SST forced interannual variations from the observed SST run are discussed in § 4. The interannual variability due to internal dynamics is also examined in this section. The amplitude of interannual variations of Indian summer monsoon indices due to internal dynamics is compared to the amplitude of forced variability. The nature of the low frequency variability in the climatological SST run is also investigated. It is discovered that the model atmosphere exhibits an internal quasi-biennial oscillation. Possible physical mechanisms for the origin of this biennial oscillation are discussed in § 5. A summary of the results is presented in § 6. The implications of our results to the predictability of Indian summer monsoon are also discussed here.

2. The model and experiments

The atmospheric model used in this study is a version of the Geophysical Fluid Dynamics Laboratory (GFDL) climate model. It is a spectral model where the dynamic computations are performed using so called spectral element method in which the distribution of a predicted variable is represented by a series of spherical harmonics and grid points. This version has a rhomboidal 30 horizontal resolution (30 zonal waves and 30 associated Legendre functions; approximately 3.75° longitude by 2.25° latitude resolution) and 14 unevenly spaced sigma levels in the vertical (R30L14 version). The vertical derivatives in the prognostic equations are computed by a centered, second order finite difference scheme. The lower resolution version (R15L09) described by Gordon and Stern [18] has been used in many climate studies [19, 20]. The physical processes and their parameterizations included in this version of the model are briefly described below.

The distribution of insolation at the top of the model atmosphere is prescribed. It varies seasonally but does not have a diurnal cycle. The solar constant is $1365\,\mathrm{W/m^2}$ and the mixing ratio of carbon dioxide is taken as 345 ppm. Solar radiation is computed using method similar to that used by Lacis and Hansen [21] except that the bulk optical properties of clouds such as reflectivity and absorptivity are prescribed. It includes multiple reflections by clouds but does not account for aerosols. Terrestrial long wave radiation is computed following the method described by Rodgers and Walshaw [22] and as modified by Stone and Manabe [23]. It includes the effect of the water vapour continuum. Ozone is prescribed as a function of latitude, height and season based on data from Hering and Borden [24].

Precipitation is computed following the 'moist convective adjustment scheme' described by Manabe *et al* [25]. Soil moisture is computed by the so called bucket method [26]. Over the continents surface temperature is computed from the condition of heat balance at the surface assuming a zero heat capacity. The model also include a gravity wave drag formulation developed by Hayashi and described by Broccoli and Manabe [27]. A simple mixing length theory for vertical diffusion of momentum, heat and moisture is also included. Prediction of cloud cover is done following the scheme used by Wetherald and Manabe [28].

Three multi-year integrations are carried out with the model. The three experiments are briefly described below.

OBS–SST run: In this run the observed monthly mean global SST [29] is prescribed. Monthly mean values are interpolated from one month to another to produce the required daily values. The soil moisture and snow cover are predicted. The model was integrated for 15 years with observed SST from January 1979 through December 1993 as boundary conditions. The model was initially spun up with mean seasonal cycle SST for five years from a resting isothermal state. The low frequency interannual variations of the observed SST associated with the El Nino and southern oscillation (ENSO) force low frequency interannual variations in the atmosphere in this run.

CLI–SST run: In this run mean seasonal cycle global SST was prescribed which was repeated every year. The soil moisture and snow cover are predicted. The mean seasonal cycle of SST was taken from Levitus [30]. The climatological sea-ice used in this run was taken from Walsh and Johnson [31] and Zwally *et al* [32]. The initial conditions for this run consisted of a resting isothermal atmosphere. The model was integrated for 40 years and the last 20 years are considered here. The model atmosphere was found to reach a quasi-equilibrium after about 15 years initial integration. In this run, except for the annual cycle forcing, there is no external interannual forcing. Internal dynamics and feedback with surface hydrological processes could still give rise to some interannual variability.

CSST–SM run: Both the seasonal cycle of SST and the seasonal cycle of soil moisture are prescribed in this run. The mean seasonal cycle of SST used was same as in the CLI–SST run. The seasonal cycle of soil moisture at each grid point were constructed from the 15 year simulations of the OBS–SST run. The snow cover feedback is also eliminated from the model. This integration was carried out for 18 years. In this run, even the surface hydrological feedback is eliminated. Therefore, any interannual variability in this run must arise purely due to internal dynamics in the presence of the annual cycle forcing. Monthly mean output is analyzed from all the runs.

3. Mean monsoon simulated by the model

The Asian summer monsoon has some unique characteristic features such as the low level westerly jet over the north Arabian sea and the easterly jet in the upper troposphere. The maximum strength of the easterly jet is at around 150 mb with its center shifted to the east

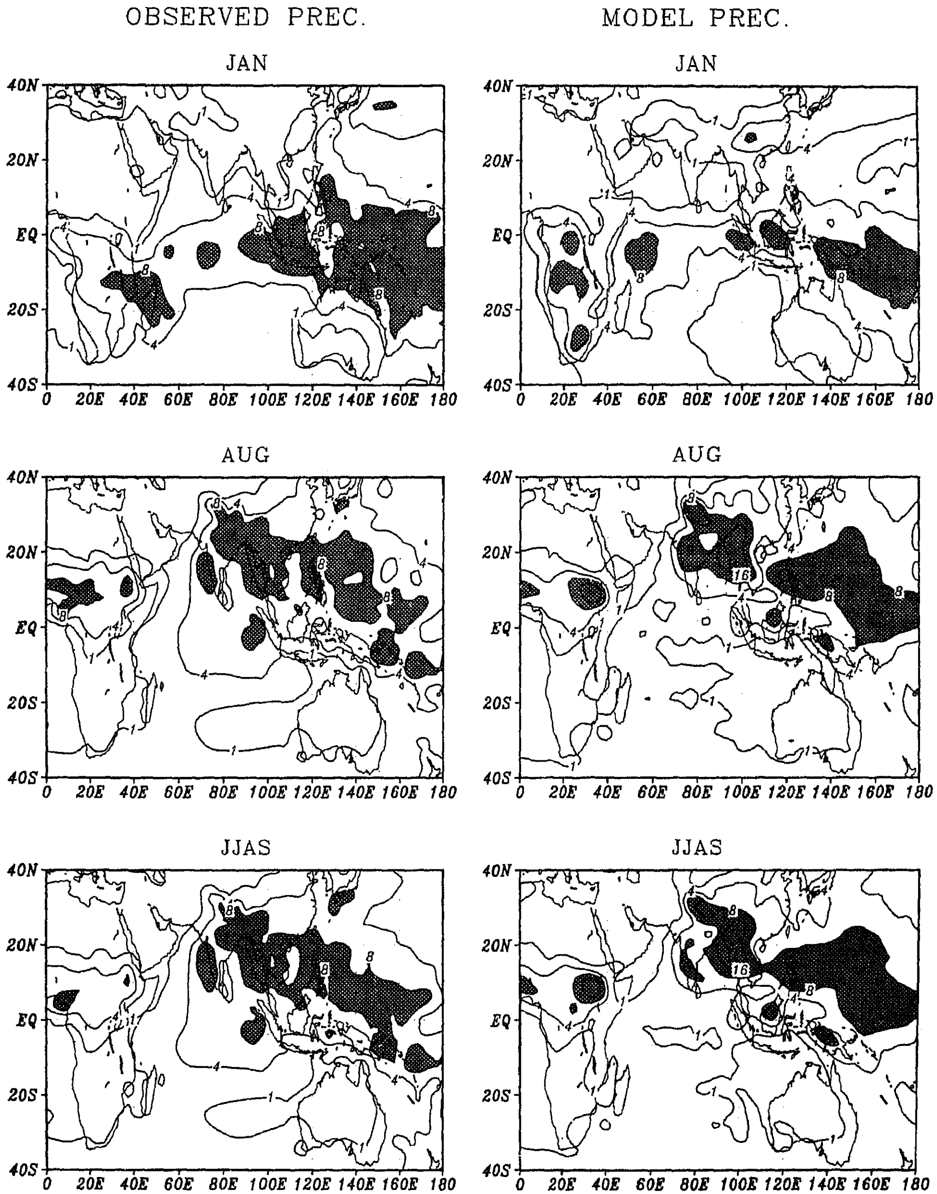


Figure 1. Mean January (JAN), August (AUG) and June–September (JJAS) precipitation. The mean derived from the OBS–SST run (right panel) is compared with observations ([33], left panels). Contours are 1, 4, 8, 16, 32 mm/day. Precipitation greater than 8 mm/day is shaded.

of the low level westerly jet center by about 10° longitude. In addition, the movement of the mean position of the ITCZ to the Indian continent and relatively sudden 'onset' and 'retreat' of the monsoon constitute other characteristic features of the Indian multi-year simulations with the same model. Since the annual cycle of the SST is nearly identical and the annual cycle of the solar forcing is identical in all three experiments, the mean annual cycle derived from the three runs should also be nearly identical. It is indeed found to be so (not shown). In figure 1, we show the precipitation climatology for January, August and June–September mean from the OBS–SST run and compare them with observations [33]. During winter (January), the position of the ITCZ is correctly simulated. However, the model simulates a smaller area of heavy precipitation (> 8 mm/day) over Indonesia than observed. During summer (August), the ITCZ does move to a position over the Indian continent. The maximum over the northern Bay of Bengal is well simulated. The precipitation over most of Indian continent is reasonably well simulated. The maximum off the western coast is not well simulated. This is related to the model's resolution being insufficient to resolve the Western Ghat mountains adequately. The seasonal mean (June–September) precipitation is also shown here as in our discussion of interannual variability we shall be referring to seasonal mean anomalies. It is noted that the seasonal mean precipitation is also well simulated by the model. The simulated precipitation over central India is slightly weaker than observed. Considering the low resolution of the model and the rather simple parameterization of cumulus convection employed by the model, the simulation of the mean precipitation is realistic. The model, however, underestimates the weaker second precipitation zone over the equatorial Indian ocean during northern summer.

We have also examined the simulation of the Somali jet, the easterly jet in the upper troposphere and the onset and retreat of monsoon by the model. It is found the model does a reasonably good job of simulating all these features.

4. Interannual variability: SST forced versus internal

In this section we discuss the interannual variability of some monsoon indices simulated by the model in the OBS–SST run. Figure 2 shows the interannual variability of the June–September (JJAS) normalized precipitation anomalies. The top panel contains the observed all India rainfall anomalies. 1982 and 1987 were deficient rainfall years while 1983 and 1988 were excess rainfall years. The lower panel shows simulated precipitation averaged only over 'Indian region' (70°–95° E, 10°–30° N, including the oceanic region included in this box). The model captures the variability during 1987 and 1988 but fails to do so during 1982. In general the model's skill in simulating the observed interannual variability of the Indian summer monsoon is rather modest. Out of the 15 years the sign of the interannual anomaly is correctly simulated in 9 years.

In order to estimate how much of the interannual variability of the monsoon discussed above is due to SST forcing and how much may be due to internal dynamics, we compare the interannual variability among the OBS–SST and CLI–SST runs. Figure 3 shows the precipitation averaged over the 'monsoon region' and the 'broad scale monsoon shear index'. The 'broad scale monsoon shear index' is defined as the zonal wind shear (U at 170 mb – U at 860 mb) averaged over EQ-20° N and 40°–100° E [11]. Both indices are

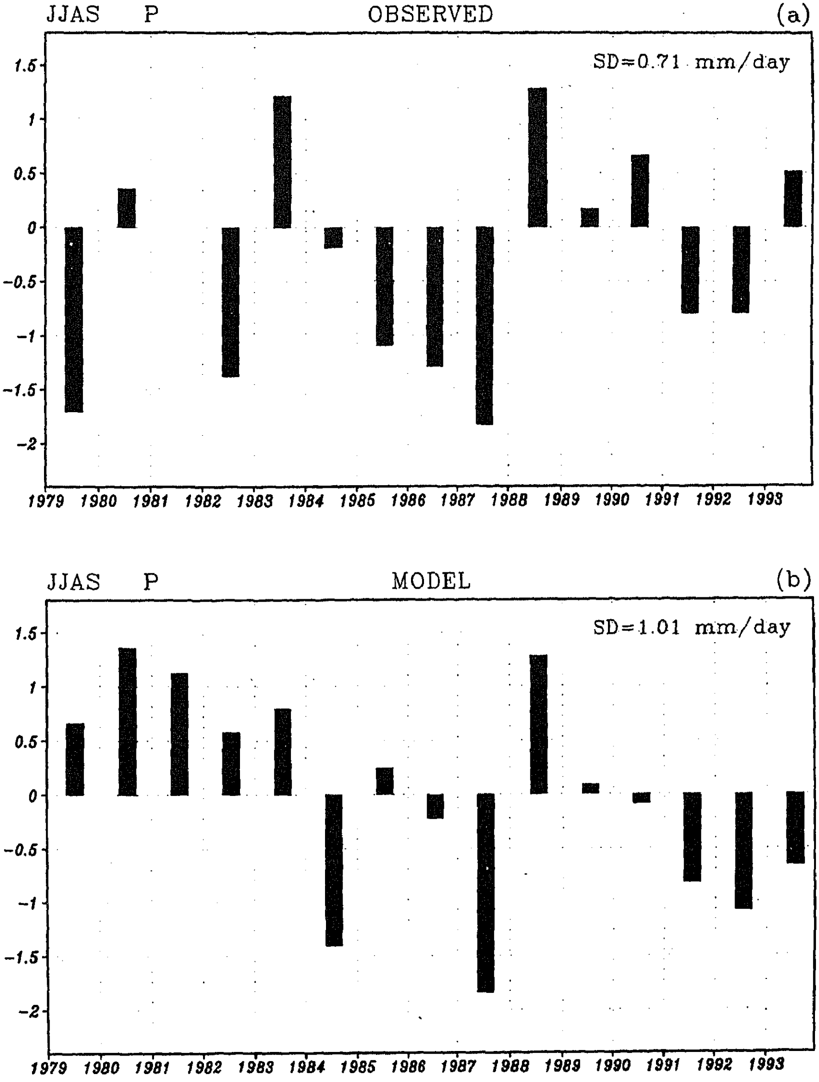


Figure 2. Comparison of simulated interannual variability of the seasonal mean monsoon (June–September) precipitation with observations. (**a**) Observed 'All-India' precipitation anomalies, (**b**) the simulated precipitation anomalies averaged over India and the Bay of Bengal (70°–95° E 10°–30° N). In both cases the precipitation anomalies are normalized with their own standard deviation.

averaged over the summer monsoon season, June–September. An examination of the interannual variability of the precipitation index shows that, although the observed SST tends to enhance the variability, the internal dynamics alone produces considerable interannual variability. The largest precipitation anomalies in the OBS–SST run occur either during strong warm or cold events in the Pacific (1987, 1988). The amplitude of the interannual variability in the remaining 'normal' years is quite close to the interannual

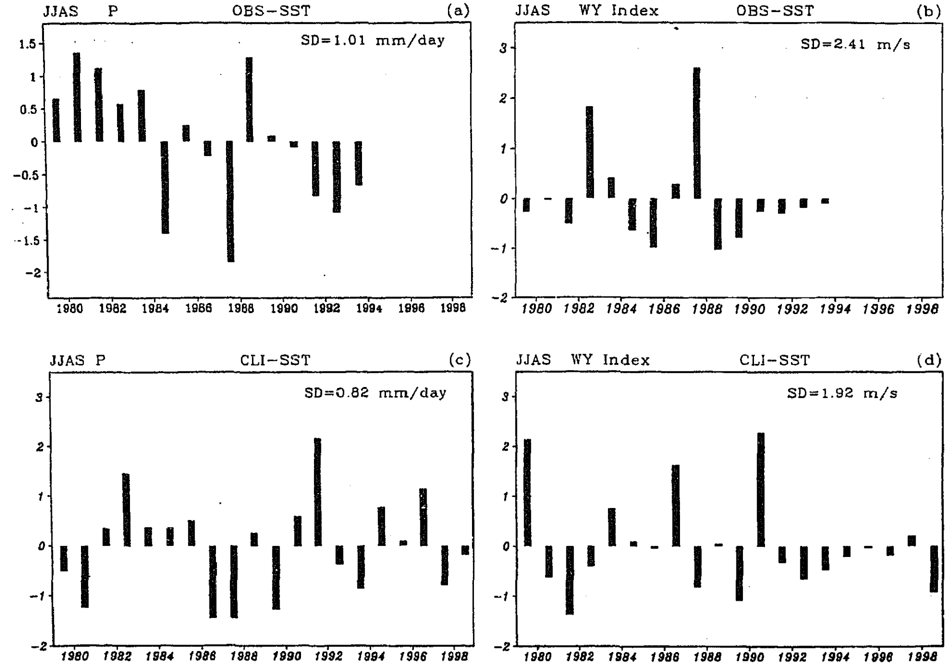


Figure 3. Comparison of interannual variability of two summer monsoon indices between OBS–SST and CLI–SST runs. Left panels correspond to JJAS averaged precipitation (mm/day) averaged over Bay of Bengal (70°–95° E, 10°–30° N), right panels correspond to JJAS averaged 'broad scale monsoon shear index' in m/s. The interannual standard deviations are noted in each panel. Results of 20 years of simulation are shown for the CLI–SST runs. The year labeling for the CLI–SST run are arbitrary and has been made to match with that of the OBS–SST run for the first 15 years. The last five years of the plot for OBS–SST run is left blank.

variability in the CLI–SST run. This indicates that strong cold (warm) events in the Pacific are associated with increased (decreased) precipitation over the monsoon region. However, if the SST anomalies are not very large during the summer season, the internal variability could dominate the variability of the monsoon precipitation. This is consistent with the predictability studies of Brankovic and Palmer [34]. The simulation of the 'broad scale monsoon shear index' also shows that the amplitude of the interannual variability generated purely by internal dynamics (CLI–SST) is, comparable to that forced by the slowly varying SST boundary forcing (OBS–SST). In the OBS–SST run the entire variability is dominated by the 1987 event. The interannual variability produced by the internal dynamics is often larger than that in the SST boundary forced run.

4.1 *A Quasi-biennial internal oscillation*

From the above discussions, it is clear that the model atmosphere can generate significant interannual variability even in the absence of any external forcing with interannual periods. What then is responsible for these interannual variations? Are there significant natural oscillations of the system or are they simply part of a red-noise process? To get

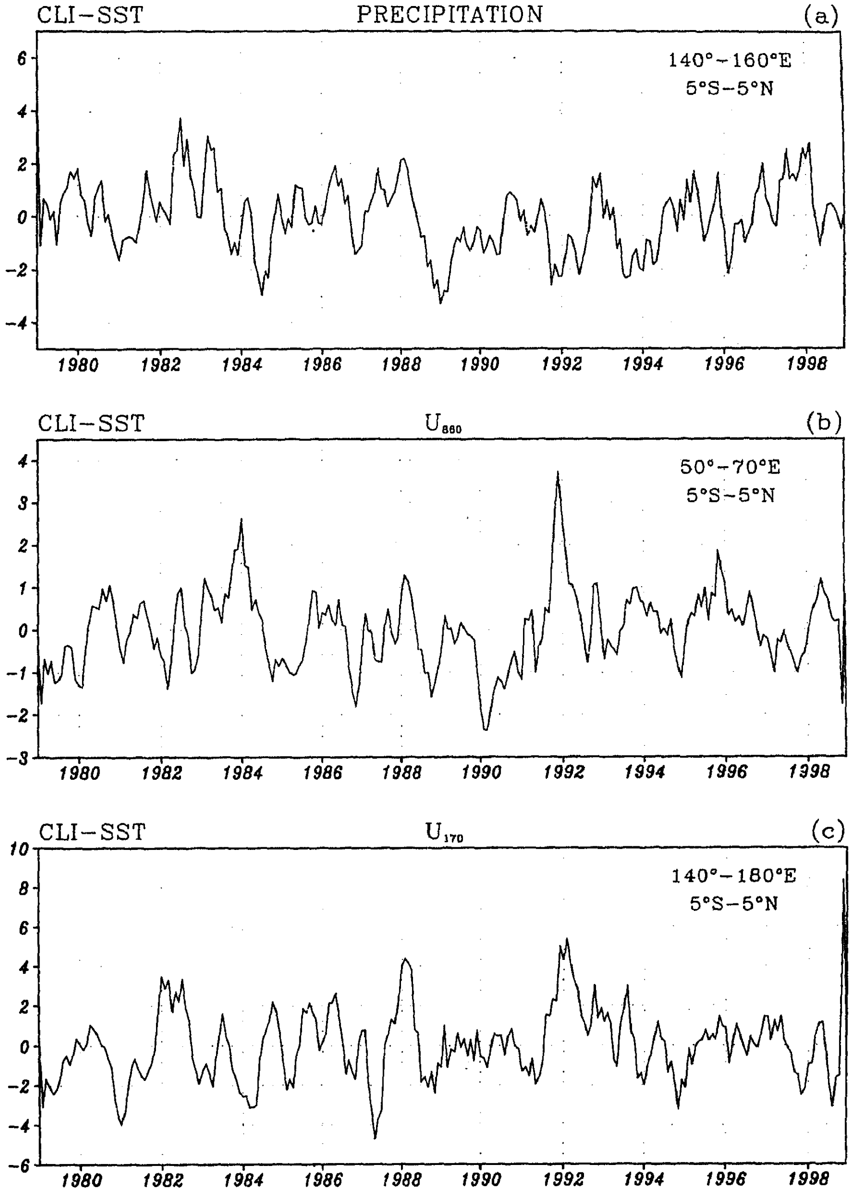


Figure 4. The interannual variability in the CLI–SST simulation. Time series of 5 month running mean equatorial (5° S–5° N) precipitation P (mm/day, 140°–160° E) and zonal wind at 170 mb (140°–180° E) and 860 mb (50°–70° E) (m/s).

some idea regarding the temporal variability of the low frequency internal oscillations, we show 5 month running means of three different variables averaged over three different regions from the CLI-SST run in figure 4. The precipitation is averaged over 5° S–5° N and 140°–160° E. The zonal winds at 170 mb and 860 mb are averaged over 5° S–5° N

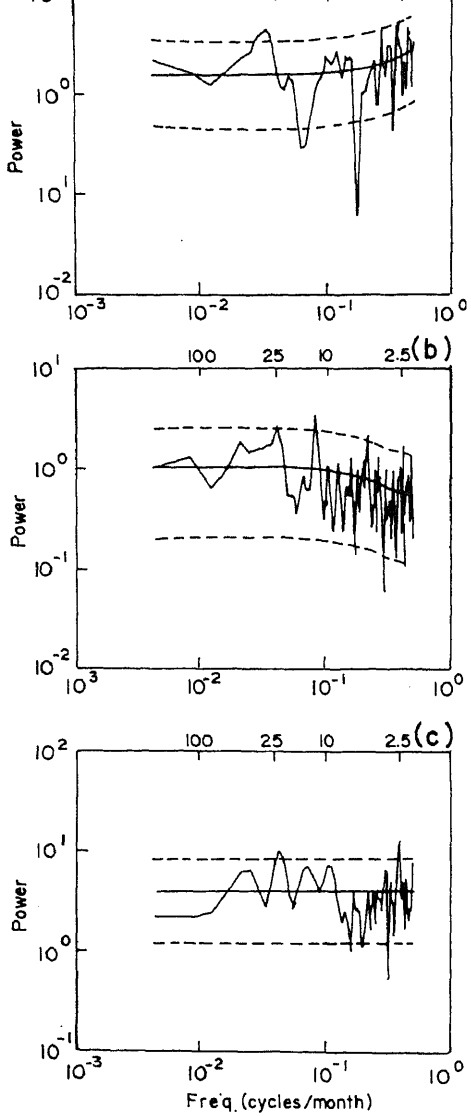


Figure 5. Power spectra of un-filtered monthly mean precipitation and zonal winds averaged over same regions shown in figure 4, for the CLI–SST run. (**a**) Precipitation (5° S–5° N, 140°–160° E), (**b**) zonal wind at 860 mb, (5° S–5° N, 50°–70° E), (**c**) zonal wind at 170 mb (5° S–5° N, 140°–180° E). The theoretical red noise spectrum based on lag 1 auto-correlation (thick solid) and the 95% confidence limits are also shown.

and 140°–180° and 50°–70° E respectively. It is clear that all the three variables show significant low frequency variability. In particular the precipitation time series tends to indicate a quasi-biennial variability. Quasi-biennial variations can be also be detected in the wind time series. This preliminary examination indicates that the model atmosphere has a quasi-biennial internal oscillation. A spectrum analysis of the original monthly mean time series confirms this (figure 5). The 95% confidence limits in figure 5 are calculated from the lag 1 auto correlations of the original time series. It is noteworthy that all three time series have a statistically significant peak around a biennial periodicity. We must add here that the three regions shown here have been selected after some trial and error. This is because the biennial variability has regional scale spatial structure. Therefore, if we average the anomalies over some arbitrary regions, there is a danger of canceling positive and negative anomalies leading to an insignificant signal.

5. Possible mechanisms

Air-sea coupling, air-land surface coupling [35–37] and interaction between nonlinear intra-seasonal oscillations and the annual cycle [38] have been invoked to explain the origin of the observed biennial component of the interannual variability in the Indian and Pacific region. As climatological SST is prescribed in our run, ocean-atmosphere coupling could not be responsible for our biennial oscillation. The other possibility is the atmosphere and land-surface feedback as envisaged by Meehl [37]. Qualitatively, this feedback may be described as follows. The strength of the monsoon circulation is related to the north-south temperature gradient over the region. A strong monsoon is associated with a strong north-south temperature gradient and results in excess precipitation and soil moisture over land. If the enhanced soil moisture persists through the next three seasons, land would be wetter than normal prior to the next monsoon. Evaporation and enhanced latent heat flux would result in cooler land temperature and weaker north-south temperature gradient leading to a weaker monsoon. A weak monsoon this year would be associated with below normal precipitation which through the same arguments would lead to a strong north-south temperature gradient and a strong monsoon next year. In this manner the cycle could be repeated every two years. While this mechanism could work in principle, there is no convincing evidence that soil moisture anomalies persist for three or more seasons. In order to determine whether this feedback is at work in producing the biennial oscillation in our model atmosphere, we carried out an additional multi-year run in which in addition to the annual cycle of SST, the annual cycle of the soil moisture and snow cover were also prescribed (CSST–SM). This experiment was conducted for 18 years. We then carried out time series and spectrum analysis of various monthly mean fields as before. It is found that this run also contains a significant biennial signal as a dominant part of its interannual variability. The spectra of un-filtered monthly mean precipitation averaged over two locations are shown in figure 6. It is clear that both the time series show a rather strong and significant peak with a period around two years. Therefore, it appears that the biennial oscillation of our model atmosphere does not owe its origin to a ground hydrology feedback as envisaged by Meehl [37].

This leaves the possibility of producing a biennial signal through interaction between atmospheric high frequency transients (synoptic and intraseasonal oscillations) and the annual cycle as a candidate for the biennial oscillation in our model atmosphere. Lorenz [39] first addressed whether the nonlinear weather events (synoptic systems) in the middle latitude modulated by the annual cycle forcing could give rise to any significant

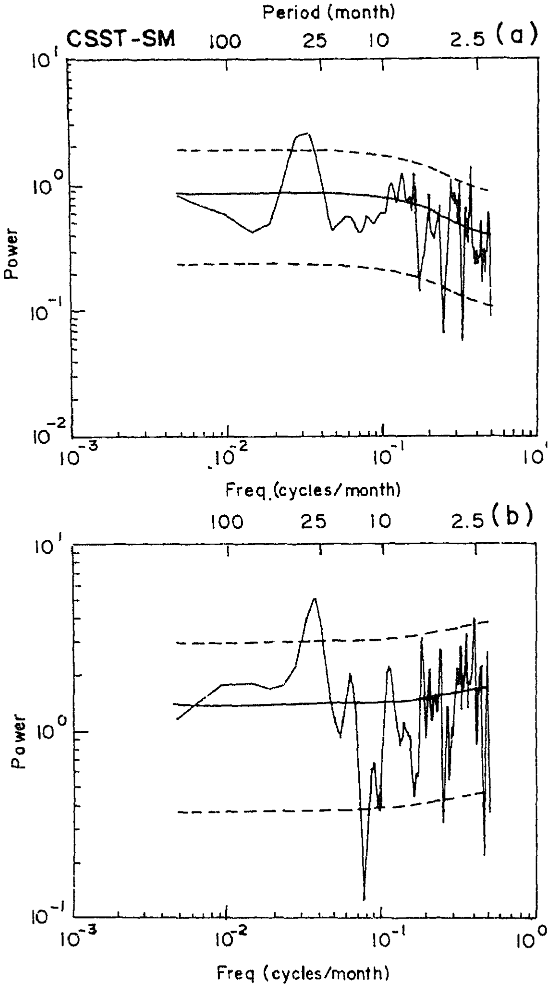


Figure 6. Power spectra of precipitation averaged between 5° S–5° N and over two longitudinal belts (**a**) (70°–95° E), and (**b**) (120°–140° E) from CSST–SM simulation. Other conventions are same as in figure 5.

interannual variability. He showed that due to the intransitive nature of the nonlinear high frequency component, a periodic forcing can give rise to some interannual variability. However, there was no suggestion of a preferential selection of the biennial oscillation. In contrast to the mid latitude, in the tropics the day to day weather fluctuations are less energetic but the intraseasonal oscillations such as the active break phases of the monsoon are quite energetic. Here we show that the modulation of the nonlinear atmospheric ‘intraseasonal oscillations’ by the annual cycle could result in an atmospheric biennial mode. To demonstrate this, we choose the following prototype nonlinear system [39]

$$\dot{X} = -Y^2 - Z^2 - aX + aF, \qquad (1)$$

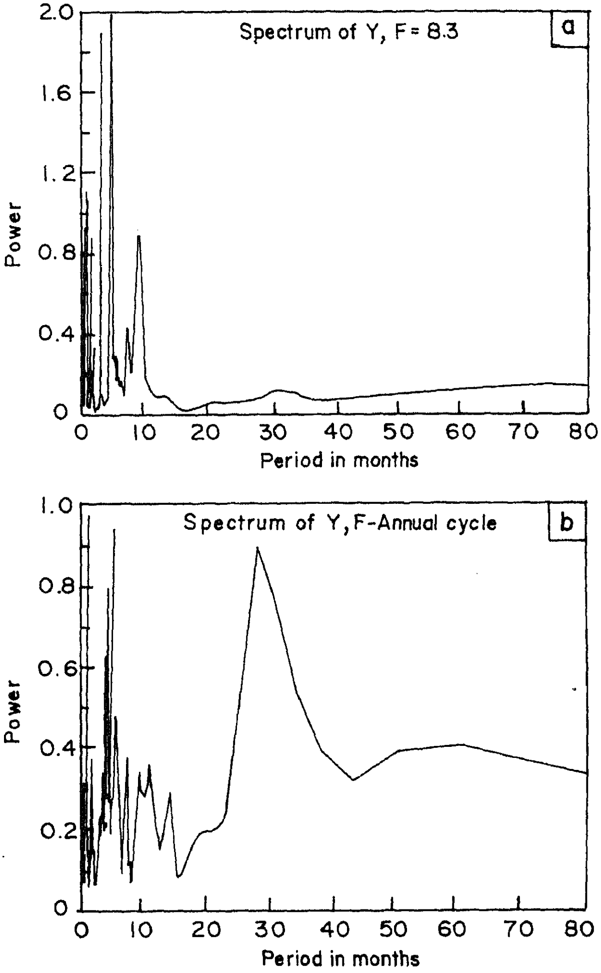


Figure 7. Spectrum of daily Y with $c = 0.5$. (**a**) for a constant $F = 8.3$ and (**b**) annual cycle of $F = 7.0 + 2.0 \cos \pi t/1$ year.

$$\dot{Y} = XY - bXZ - cY + G, \tag{2}$$

$$\dot{Z} = bXY + XZ - cZ, \tag{3}$$

where X may be considered as the zonally symmetric component while Y and Z may be interpreted as non-zonal components. F is the zonally symmetric forcing (e.g., solar forcing) while G is the non-zonal forcing arising out of land-ocean contrasts. Values of the parameters a and b representing dissipation and strength of eddy-mean flow interaction used by Lorenz are retained ($a = 0.25$ and $b = 4.0$). Also, $G = 1.0$. Original equations of Lorenz and the parameters are all scaled by a factor c. For $c = 1.0$, the spectrum of the nonlinear system within the chaotic regime shows maximum power around synoptic frequencies (around one week). With $c = 0.5$, however, the preferential period in the chaotic regime shifts to intraseasonal range (3 to 5 months). In figure 7(a),

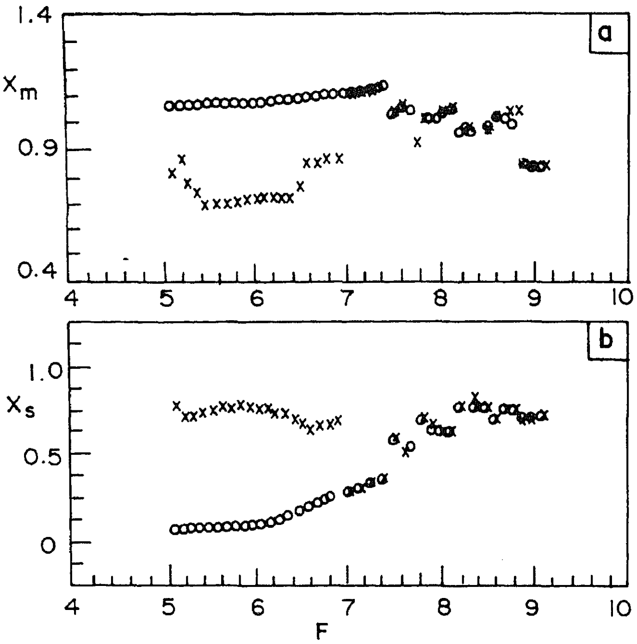


Figure 8. Some gross features of the attractors for different values of constant forcing, F. (**a**) Long term mean, X_m and (**b**) standard deviation, X_s. The two symbols represent values obtained with two different initial conditions for each forcing.

the spectrum of daily Y for a fixed $F = 8.3$ is shown based on the last 100 years of a 150 year integration. Next we introduced an annual cycle forcing with $F = 7.0+2.0\cos\pi t/\tau$, where τ is one year. Again, the spectrum of daily Y is shown based on the last 100 yeas of a 150 year integration. The modulation of the intraseasonal oscillations by the annual cycle clearly gives rise to a significant biennial oscillation (figure 7b). If, however, $c = 1.0$ and the nonlinear oscillations are primarily confined to the synoptic range, the preferential excitation of the biennial mode is very weak (not shown).

How does periodic variation of forcing for the atmosphere with a period of 1 year generate responses at all low frequencies? This was discussed by Lorenz [39] in detail. To understand how the slowly varying forcing introduces the low-frequency response, we need to examine the nature of variations of the nonlinear system (1)–(3) for different values of the steady forcing. In figure 8, we show some gross characteristics of the attractors (long-term mean and standard deviation) of the system (1)–(3) for a range of values of F. The two symbols represent attractors attained from two different initial conditions. The system has periodic attractors for F between 5.0 and 7.5. For F between 7.5 and 8.8, the system has a chaotic attractor. Beyond $F = 8.8$, again the system goes to a periodic regime. We note that for F between 5.0 and 7.0, the system has at least two periodic orbits, one with high amplitude and low mean, while the other with low amplitude and higher mean. Within the chaotic regime too, the system tends to have two preferred regimes, one with large amplitude oscillations with low mean (active regime) and another with small amplitude oscillations with high mean (weak regime). This is seen in figure 9 where the frequency distribution of mean and standard deviation calculated

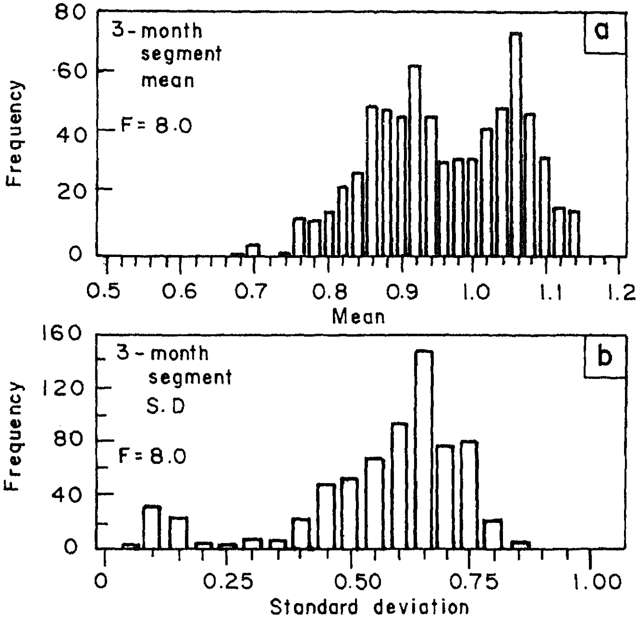


Figure 9. Frequency distribution of (**a**) mean and (**b**) standard deviation of three month segments for $F = 8.0$ over a period of 184 years.

over three month segments over a period of 184 year with a constant $F = 8.0$ (chaotic regime) is shown. Both the standard deviation and the short term mean show the tendency for two preferred regimes.

With this background, we may try to understand the response of the nonlinear system when the forcing varies annually between two extremes, say between $F = 9$ and $F = 5$, the system reaches a chaotic attractor that consists of a strong and a weak regime. The two regimes resemble the orbits of the multiple periodic attractors of constant forcings discussed earlier. Whether the system would tend to visit the weak or the strong regime in a given period of the forcing depends on the previous history of state (or initial conditions, in a loose manner). At many of these values of F, when they are held constant, the system may possess multiple periodic attractors that depend on the initial conditions. But when the forcing is varying because of chaos, the system reaches a different state on the same day (or same value of t) of every year. This difference determines the difference in the subsequent behaviour of the system-whether to visit the strong or the weak regime. This essentially is why the annual cycle itself is capable of introducing some interannual variability. It appears that in one year the system tends to reside more in the low amplitude-high mean state while in another year it tends to reside in the high amplitude low mean state giving rise to a biennial variability.

6. Summary of results and implications for monsoon predictability

The interannual variability of the Indian summer monsoon simulated by three multi-year simulations with the GFDL R30L14 model is studied. The monsoon climatology of the model is realistic including the position and strength of the low level westerly jet, upper level easterly jet and precipitation over continental India. The skill of simulation of interannual variability of the Indian summer precipitation by the GFDL climate model is only modest. Most general circulation models are found to be deficient in simulating the observed interannual variability of the Indian summer monsoon precipitation [15]. The rather poor predictability of Indian summer monsoon precipitation is shown to be due to the fact that the interannual variability forced by slowly varying SST over the Indian region is comparable or even weaker than that due to internal dynamics.

Further examination of the nature of the interannual variability in the CLI–SST run reveals that much of the interannual variability of the model is due to a significant biennial oscillation of the model atmosphere. The CSST–SM experiment was carried out to address whether the biennial oscillation could have resulted from a soil moisture-atmosphere feedback or whether it was internal to the atmosphere. The fact that a significant biennial oscillation is seen even in the fixed soil moisture run rules out the possibility that ground hydrology feedback is responsible for the biennial oscillation. This leaves the possibility of producing a biennial signal through interaction between atmospheric high frequency transients (synoptic and intra-seasonal oscillations) and the annual cycle as proposed by Goswami [38], as a candidate for the biennial oscillation in our model atmosphere. This was demonstrated with a simple paradigm model of the tropical atmosphere. The annual cycle may be considered as a slowly varying forcing for the nonlinear atmospheric high frequency oscillations. As the nonlinear atmospheric oscillations go through a chaotic regime during parts of the slowly varying periodic forcing, it can result in a low frequency tail in the atmospheric oscillations.

The biennial internal oscillation is likely to be a major limiting factor for prediction of seasonal monsoon precipitation. The biennial mode has largest amplitude around the precipitation zone over the Indonesian region. As the SST forced response is relatively weak over the Indian region, and the biennial internal mode has large amplitude, it may dominate the interannual variability. Although the amplitude of the biennial mode is large over northern Australia, the SST forced response is also large. As a result the biennial mode cannot dominate the interannual variability over northern Australia. Over the Sahel region in Africa the amplitude of the biennial oscillation is very weak. Therefore, the interannual variations over Sahel are again primarily governed by SST variations. These conclusions from our analysis are consistent with conclusions that the seasonal predictions of precipitation over the Sahel [17] and northern Australia based on SST can be made with much greater confidence than can be done over the Indian monsoon region [34]. The discovery of a quasi biennial internal oscillation of the tropical atmosphere limiting the predictability of the Indian monsoon needs to be reexamined with other general circulation models.

Acknowledgments

The author is indebted to Suki Manabe without whose time, help and encouragement this work could not have been completed. He made the results of the already existing CLI–SST experiment with the same model available and helped in extending the OBS–SST run for the latest five years. He also got the new CSST–SM run conducted for this study. A number of insightful discussions with him helped the author design the experiments. This work is partially supported by a grant from Department of Science and Technology, Government of India. Some of the calculations were carried out at the Supercomputer Education and Research Centre of the Indian Institute of Science.

References

[1] E N Lorenz, *Tellus* **17**, 321 (1965)
[2] J G Charney and J Shukla, *Predictability of monsoons, Monsoon Dynamics* edited by J Lighthill and R P Pearce (Cambridge University Press, Cambridge, 1981) vol. 99
[3] B N Goswami, V Krishnamurthy and N H Saji, *Mon. Weather Rev.* **123**, 1677 (1995)
[4] D J Hahn and J Shukla, *J. Atmos. Sci.* **33**, 2461 (1976)
[5] R R Dickson, *J. Climate Appl. Meter.* **23**, 171 (1984)
[6] A D Vernekar, J Zhou and J Shukla, *J. Climate* **8**, 248 (1995)
[7] J Shukla and Y Mintz, *Science* **214**, 1498 (1982)
[8] G T Walker, *Mem. Indian Meteor. Dept.* **23**, 23 (1924)
[9] J Shukla and D A Paolino, *Mon. Weather Rev.* **111**, 1830 (1983)
[10] T Yasunari, *Bull. Am. Meteor. Soc.* **72**, 1331 (1991)
[11] P J Webster and S Yang, *Quart. J. Roy. Meteor. Soc.* **118**, 877 (1992)
[12] T N Palmer, C Brankovic, P Viterbo and M J Miller, *J. Climate* **5**, 399 (1992)
[13] F W Zwiers, *J. Climate* **6**, 470 (1993)
[14] T C Chen and M C Yen, *J. Climate* **8**, 1403 (1994)
[15] K Sperber and T N Palmer, Atmospheric Model Inter-comparison Project: Monsoon Simulations, *Proc. Int. Conf. on Monsoon Variability and Prediction*, Trieste, May 9–13, 1994, WMO/TD-619, 601
[16] M Fennessy and J Shukla, Simulation and predictability of monsoons, *Proc. Int. Conf. on Monsoon Variability and Prediction*, Trieste, May 9–13, 1994 WMO/TD-619, 567
[17] T N Palmer and D Anderson, *Quart. J. Roy. Meteor. Soc.* **120**, 755 (1994)
[18] C T Gordon and W Stern, *Mon. Weather Rev.* **110**, 625 (1982)
[19] N C Lau and M Nath, *J. Climate* **7**, 1184 (1994)
[20] S Manabe and R J Stouffer, *J. Climate* **7**, 5 (1994)
[21] A Lacis and J Hensen, *J. Atmos. Sci.* **31**, 118 (1974)
[22] C D Rodgers and C D Walshaw, *Quart. J. Roy. Meteor. Soc.* **92**, 67 (1966)
[23] H M Stone and S Manabe, *Mon. Weather Rev.* **96**, 735 (1968)
[24] W S Hering and T R Borden Jr., Mean measurements of ozone density over north America, 1963–1964. *Environmental Research Papers*, Report 162 (US Air Force Res. Lab., Mass., 1965)
[25] S Manabe, J Smagorinsky and R F Strickler, *Mon. Weather Rev.* **93**, 769 (1965)
[26] S Manabe, *Mon. Weatehr Rev.*, **97**, 739 (1969)
[27] A J Broccoli and S Manabe, *J. Climate* **5**, 1181 (1992)
[28] R W Wetherald and S Manabe, *J. Atmos. Sci.* **45**, 1397 (1988)
[29] R Reynolds, *J. Climate* **1**, 75 (1988)
[30] C Levitus, Climatological atlas of the world ocean. *NOAA Professional Paper No. 13* (U.S. Department of Commerce, Rockville, MD, 1982)
[31] J E Walsh and C M Johnson, *J. Phys. Oceanogr.* **9**, 580 (1979)
[32] H M Zwally, J C Comiso, C L Parkinson, W J Campbell, F D Carsey and P Gloersen, *Antarctic sea ice 1973–1976: satellite passive microwave observations*. NASA SP-459, (NTIS N84-10718/4), (1983)
[33] D Legates and C J Willmott, *Int. J. Climatol.* **10**, 111 (1990)
[34] C Brankovic and T Palmer, Predictability of summer monsoon, *Proc. Int. Conf. on Monsoon variability and prediction*, Trieste, May 9–13, 1994, WMO/TD-619, 629
[35] G A Meehl, *Mon. Weather Rev.*, **115**, 27 (1987)

[36] G A Meehl, *J. Climate* **6**, 31 (1993)
[37] G A Meehl, *J. Climate* **7**, 1033 (1994)
[38] B N Goswami, *J. Climate* **8**, 524 (1995)
[39] E N Lorenz, *Tellus* **42A**, 378 (1990)

—Journal of
physics

February 1997
pp. 737–748

Some recent applications of cell dynamical modelling to phase ordering dynamics

SANJAY PURI
School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Abstract. We study experimentally relevant effects in phase ordering dynamics using Cell Dynamical System (CDS) models. In particular, we present representative numerical results for phase ordering in random magnets and phase separation in binary fluids.

Keywords. Cell dynamical modelling; phase ordering dynamics; disordered systems.

PACS Nos 64.75; 64.60

1. Introduction

There has been much recent interest in the problem of phase ordering dynamics, viz., the temporal evolution of a homogeneous two-phase mixture which has been rendered thermodynamically unstable by a sudden quench below its critical temperature [1]. Typically, the evolving system separates into domains which are rich in one or the other constituent of the mixture. These domains coarsen with time and are characterised by a single length scale $R(t)$ at time t (for isotropic systems). A direct consequence of the existence of a unique length scale is the so-called dynamical scaling of the time-dependent correlation function $g(\vec{r})$, i.e., $g(\vec{r}) = G(r/R(t))$, where $\vec{r}$ is the spatial variable and $G(x)$ is a master function, which is independent of time [2]. The nature of the domain growth law; and the scaling form of the correlation function are of great experimental relevance in disciplines as diverse as physics, chemistry, materials science and metallurgy. They depend critically on a variety of factors as follows:

(a) whether the evolution is characterised by a nonconserved order parameter (e.g., ordering of a ferromagnet) or conserved order parameter (e.g., segregation of a binary alloy);
(b) whether or not hydrodynamic effects are relevant;
(c) the nature of disorders (quenched or annealed) present in the phase ordering system, etc.

Theoretical descriptions of phase ordering dynamics are usually in terms of nonlinear partial differential equations, which are analytically intractable – especially in the case with conserved order parameter. Thus, it is necessary to have computationally efficient numerical models which enable us to access the asymptotic behaviour of phase ordering dynamics with limited numerical effort. Some time ago, we formulated discrete from their relevance in phase ordering dynamics, our CDS models have also served as an important example of the relevance of discrete space-time models in the modelling of nonequilibrium phenomena.

In this paper, we would like to briefly discuss two experimentally relevant problems of phase ordering dynamics, which we have recently tackled using CDS models. This paper is organised in the following fashion. In § 2, we discuss conventional models of phase ordering dynamics and also describe our CDS models. In § 3, we study the first problem of interest to us here, viz., phase ordering dynamics in systems with quenched disorder. Section 4 of this paper examines the second problem of interest to us here, viz., the role of hydrodynamic effects in phase separation dynamics. Section 5 ends this paper with a summary and discussion.

2. Conventional and cell dynamical modelling of phase ordering dynamics

At the coarsegrained level, phase ordering dynamics is conventionally modelled via stochastic partial differential equations. Typically, systems with a nonconserved order parameter (e.g., ordering of ferromagnets) are described by the time-dependent Ginzburg-Landau (TDGL) equation:

$$\frac{\partial\psi(\vec{r},t)}{\partial t} = -\frac{\delta H[\psi(\vec{r},t)]}{\delta\psi(\vec{r},t)} + \sigma(\vec{r},t), \tag{1}$$

where $\psi(\vec{r},t)$ is the order parameter of the system at point $\vec{r}$ at time t. For pure and isotropic systems, $H[\psi(\vec{r},t)]$ is usually taken to be the coarse-grained ϕ^4-free energy functional:

$$H[\psi(\vec{r},t)] = \int \mathrm{d}\vec{r}\,[-\tfrac{1}{2}\psi(\vec{r},t)^2 + \tfrac{1}{4}\psi(\vec{r},t)^4 + \tfrac{1}{2}(\vec{\nabla}\psi(\vec{r},t))^2], \tag{2}$$

where we assume that all quantities have been rescaled into appropriate dimensionless units. The Gaussian white noise $\sigma(\vec{r},t)$ in (1) has zero expectation value and satisfies the usual fluctuation-dissipation relation. In the classification scheme of Hohenberg and Halperin [4], Eq. (1) is referred to as Model A.

Of course, real experimental systems are neither pure nor isotropic and the free-energy functional in (2) constitutes an oversimplification. For example, in ferromagnets, quenched disorder typically arises in the form of nonmagnetic impurities. At the phenomenological level, this is modelled by assigning a random spatial dependence to the coefficients of various terms in the free-energy functional of (2) [5, 6]. The distribution of this random fluctuation is usually taken to be either Gaussian or uniform. For studies of phase ordering dynamics, the precise nature of this distribution is not relevant [6].

Phase ordering dynamics in systems with a conserved order parameter but no hydrodynamic effects (e.g., segregation of a binary alloy AB) is described by the

Cahn–Hilliard–Cook (CHC) equation [7]:

$$\frac{\partial\psi(\vec{r},t)}{\partial t}=\nabla^2\left[\frac{\delta H[\psi(\vec{r},t)]}{\delta\psi(\vec{r},t)}\right]+\sigma(\vec{r},t), \tag{3}$$

where symbols have the same meaning as previously. In the case with conserved order parameter, $\int d\vec{r}\psi(\vec{r},t)$ is a constant in time, as is clear on integrating Eq. (3) over space. Eq. (3) is also referred to as Model B [4].

The simplest model which incorporates hydrodynamic effects in segregation dynamics is known as Model H [4] and consists of a conserved scalar density (the order parameter) coupled to the hydrodynamic velocity field. We will consider here a variant of Model H, with the following form in three dimensions (again in appropriate dimensionless units):

$$\begin{aligned}\frac{\partial\psi(\vec{r},t)}{\partial t}&=\nabla^2\left[\frac{\delta H[\psi(\vec{r},t)]}{\delta\psi(\vec{r},t)}\right]-\bar{\alpha}\vec{\nabla}\cdot[\psi(\vec{r},t)\vec{J}(\vec{r},t)]+\sigma(\vec{r},t),\\ \frac{\partial J_i(\vec{r},t)}{\partial t}&=\bar{\eta}\nabla^2 J_i(\vec{r},t)+\bar{\sigma}\sum_{k=1}^{3}\nabla_i\nabla_k J_k(\vec{r},t)\\ &\quad-\bar{\alpha}\psi(\vec{r},t)\nabla_i\left[\frac{\delta H[\psi(\vec{r},t)]}{\delta\psi(\vec{r},t)}\right]+\nu_i(\vec{r},t).\end{aligned} \tag{4}$$

In (4), $\psi(\vec{r},t)$ and $J_i(\vec{r},t)$ $(i=1,2,3)$ are, respectively, the order parameter and the dimensionless velocity fields as a function of dimensionless space and time. The Gaussian white noises $\sigma(\vec{r},t)$ and $\nu_i(\vec{r},t)$ $(i=1,2,3)$ are the noises for the order parameter field and the velocity field, respectively. They satisfy the appropriate fluctuation-dissipation relations. The rescaled parameters in (4) are the coupling constant $\bar{\alpha}$ ($\bar{\alpha}=0$ corresponds to the usual case of binary alloys, viz., Model B); and the transport coefficients $\bar{\eta}$ and $\bar{\sigma}$, which are the viscosities. We should remark that (4) reduces to the standard form of Model H [4] if we impose the additional constraint $\vec{\nabla}\cdot\vec{J}(\vec{r},t)=0$. However, as was pointed out by Farrell and Valls [8], this extra constraint causes considerable numerical complications and we do not impose it here.

Given the numerical effort involved in successfully simulating the stochastic partial differential equation models in Eqs. (1)–(4) or performing Monte Carlo (MC) simulations of appropriate microscopic models, it is desirable to have computationally efficient models to study the scaling regime. We have proposed CDS models of phase ordering dynamics, which are highly efficient in terms of computer time usage [3]. These models impose the space-time coarsening, implicit in conventional phenomenological models, by using a discretised space-time lattice. An order parameter variable is associated with each cell of the spatial lattice, which can be considered to be the result of coarse-graining the microscopic model. The underlying principles of our modelling have already been discussed extensively in the literature and we refer the interested reader to ref. [3]. Here, we merely confine ourselves to presenting the resultant models, which are formulated on a lattice whose cells are labelled by, say, n. We associate an order parameter value at (discrete) time t with each cell and refer to it as $\psi(t,n)$. Our CDS model for the case with nonconserved order parameter in d-dimensional space takes the following form [3]:

$$\begin{aligned}\psi(t+1,n) &= f(\psi(t,n)) + \frac{D}{2d}\Delta_D\psi(t,n)\\ &\equiv \mathcal{F}[\psi(t,n)],\end{aligned} \tag{5}$$

where $f(\psi)$ is a monotonic function with an unstable fixed point at $\psi = 0$ (which mimics the unstable homogeneous state); and two symmetrically located stable fixed points at $\psi = \pm\psi_0$ (which mimic the stable spontaneously ordered states). In (5), D is a positive constant proportional to the phenomenological diffusion constant; and Δ_D is the isotropically discretised Laplacian operator at the site n. As the local order parameter takes continuous values, this CDS model falls into the subclass of Coupled Map Lattice (CML) models [9, 10].

We can also incorporate thermal noise terms in the deterministic model of Eq. (5). However, we have demonstrated numerically that thermal noise is asymptotically irrelevant for phase ordering dynamics – at least in the context of pure systems [3]. Therefore, the deterministic model in (5) suffices if one is interested in universal asymptotic behaviour. Furthermore, we should stress that any function $f(\psi)$ with the specified properties provides a reasonable CDS model for phase ordering with a nonconserved order parameter. This is a consequence of the strong universality of the ordering process and is discussed at length in Ref. [3].

The corresponding CDS model for the case with quenched disorder, e.g., ordering in random magnets, is obtained as a simple extension of (5) by associating a random spatial dependence with the parameters of the function $f(\psi)$, so that the local fixed points are randomly distributed around $\pm\psi_0$ [6]. In our subsequent discussion, the term "amplitude of disorder" refers to the amplitude of fluctuations in the local fixed points of the function $f(\psi)$.

The corresponding CDS model for the conserved case is easy to obtain. Naively, it may be obtained analogously to the continuum case by appending an extra Laplacian operator to the chemical potential for the nonconserved case as follows [3]:

$$\psi(t+1,n) = \psi(t,n) - \frac{1}{2d}\Delta_D[\mathcal{F}[\psi(t,n)] - \psi(t,n)]. \tag{6}$$

Our procedure for obtaining CDS models can also be interpreted as a novel method of discretising the corresponding partial differential equations, which enables the use of rather coarse mesh sizes, without any loss of numerical stability [3]. As a matter of fact, we can use this interpretation to derive the CDS model equivalent to (4), i.e., for the case with conserved order parameter when hydrodynamic effects are relevant. The CDS model we obtain from (4) (again neglecting noise terms) has the following form in three dimensions [11]:

$$\begin{aligned}\psi(t+1,n) &= \psi(t,n) - \tfrac{1}{6}\Delta_D[\mathcal{F}[\psi(t,n)] - \psi(t,n)] - \alpha\vec{\nabla}_D\cdot[\psi(t,n)\vec{J}(t,n)],\\ J_i(t+1,n) &= J_i(t,n) + \eta\Delta_D J_i(t,n) + \sigma\sum_{k=1}^{3}\nabla_{D,i}\nabla_{D,k}J_k(t,n)\\ &\quad + \alpha\psi(t,n)\nabla_{D,i}[\mathcal{F}[\psi(t,n)] - \psi(t,n)],\end{aligned} \tag{7}$$

where α, η and σ are phenomenological parameters. They are analogous to the corresponding parameters in the continuum model (4). In (7), $\vec{\nabla}_D$ is the symmetrically discretised gradient operator. The values of the parameters are dictated by the requirements that the scheme be stable and that the results be physically reasonable [3, 11].

We will next describe some recent results obtained using the CDS models we have just discussed. As mentioned earlier, we focus on two experimentally relevant problems, viz., phase ordering in random magnets; and phase separation in binary fluids.

3. Phase ordering dynamics in disordered magnets

We have studied phase ordering in random magnets by implementing the disordered version of the CDS model defined by (5) on a 2-dimensional lattice of size N^2. Periodic boundary conditions were applied to both edges of the lattice. The parameter values for our simulations and the manner in which we introduce disorder in $f(\psi)$ are discussed at length in Ref. [6] and we do not replicate the discussion here.

The initial condition for a single run corresponds to small-amplitude uniformly-distributed random fluctuations about a zero background, mimicking the disordered state before the quench. The statistical quantities of interest in phase ordering dynamics are the time-dependent structure factor, which is the Fourier transform of the correlation function, and the characteristic domain length scale $R(t)$. The time-dependent structure factor is defined as

$$S(\vec{k},t) = [\langle \psi(\vec{k},t)\psi(\vec{k},t)^* \rangle], \tag{8}$$

where $\psi(\vec{k},t)$ is the Fourier transform of the order parameter (spontaneous magnetisation, in this case) with wave-vector $\vec{k}$; and $\psi(\vec{k},t)^*$ is the complex conjugate of $\psi(\vec{k},t)$. In (8), the angular brackets refer to an averaging over independent initial conditions and the square brackets refer to an averaging over disorder configurations. In the dynamical scaling regime, the structure factor has the scaling form $S(\vec{k},t) = R(t)^d F(kR(t))$, where $F(x)$ is a time-independent master function.

All numerical results presented below are obtained on 128^2 systems as averages over 20 disorder configurations, with 20 independent initial conditions for each disorder configuration. Because of the discrete nature of our lattice, the wavevectors $\vec{k}$ take discrete values $(2\pi/N)(k_x, k_y)$, where k_x and k_y range from $-N/2$ to $(N/2) - 1$. As the system of interest is isotropic, we can improve statistics by spherically averaging the vector function $S(\vec{k},t)$ to obtain a scalar function $S(k,t)$. This is the quantity we show in subsequent figures.

As mentioned earlier, the second quantity of interest is the time-dependent length scale $R(t)$. We define $R(t)$ as the reciprocal of the first moment of the scalarised structure factor, i.e., $R(t) = \langle k \rangle^{-1}$, where

$$\langle k \rangle = \frac{\int_0^{k_m} \mathrm{d}kkS(k,t)}{\int_0^{k_m} \mathrm{d}kS(k,t)}. \tag{9}$$

In (9), the upper cut-off k_m is taken to be half the magnitude of the largest wavevector lying in the Brillouin zone of the lattice. We have confirmed that the structure factor has

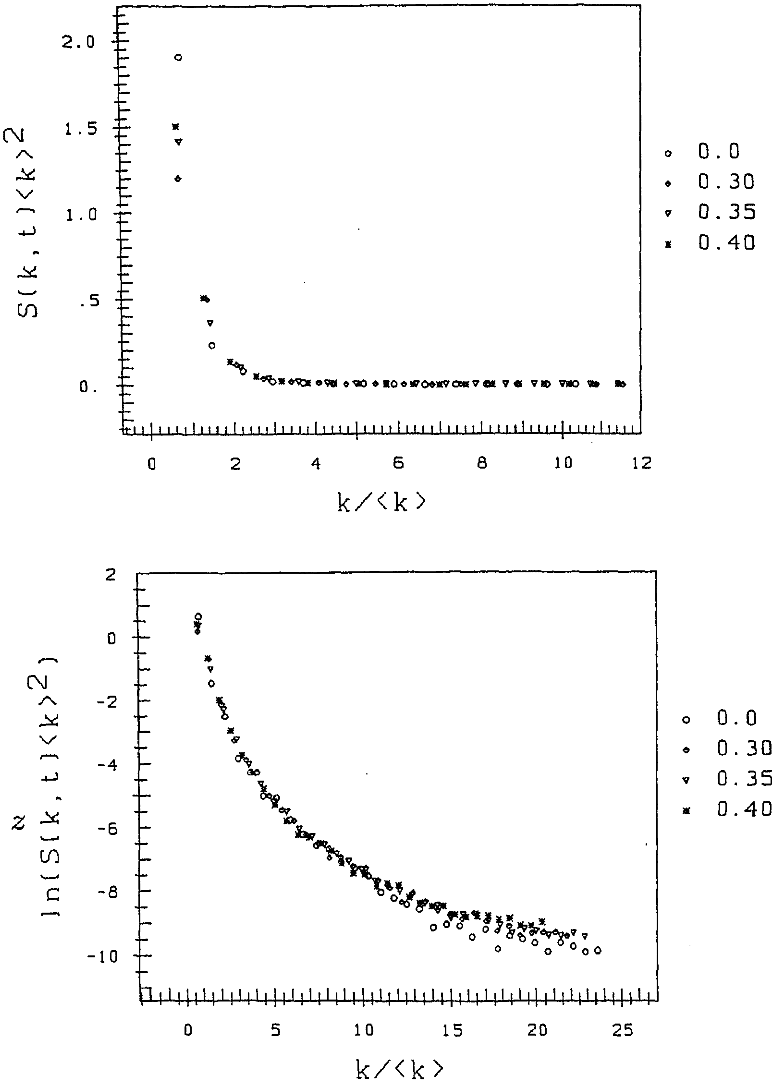


Figure 1. (a) Superposition of data for the scaled structure factor $S(k,t)\langle k\rangle^2$ vs. $k/\langle k\rangle$ obtained from CDS studies of phase ordering dynamics in random magnets. All structure factor data is obtained on 128^2 systems as an average over 20 disorder configurations, with 20 independent initial conditions for each disorder configuration. We present data for the pure case (denoted by circles); and a range of disorder amplitudes (denoted by the symbols indicated). The manner in which disorder is introduced into the CDS models is detailed in ref. [6]. The term "disorder amplitude" refers to the amplitude of random spatial fluctuations introduced into the local fixed points of the function $f(\psi)$ in Eq. (5). Data for the pure case (disorder amplitude $C = 0.0$) is from dimensionless time $t = 1000$. Data for the disordered cases ($C = 0.30, 0.35$ and 0.40) is from dimensionless time $t = 3000$. (From Ref. [6].) (b) Semi-log plot of data in figure 1(a), i.e., we plot $\ln(S(k,t)\langle k\rangle^2)$ vs. $k/\langle k\rangle$. (From ref. [6].)

decayed sufficiently upto this value of k_m and that numerical results for $R(t)$ are unchanged by considering larger values of k_m. The length scale thus defined is measured in units of the lattice spacing. We should stress that this definition of $R(t)$ is not unique and one can also define $R(t)$ by using higher moments of the structure factor or various zero-crossings of the correlation function. However, in the dynamical scaling regime, all these definitions are equivalent with the only difference being in numerical prefactors [2].

Our extensive numerical investigations of phase ordering dynamics with quenched disorder revealed the following facts. Firstly, we found that the coarsening systems exhibit dynamical scaling of the structure factor as in the pure case, indicating thereby that the presence of quenched disorder does not affect the existence of a unique domain length scale. Furthermore, we found that the functional form of the scaled structure factor appears to be independent of the amplitude of disorder and is the same as that for the pure case – at least within the rather wide range of disorder amplitudes considered. We have referred to this important property as "superuniversal" scaling of the structure factor [6]. Figure 1 exhibits the superuniversal scaling of the structure factor for phase ordering in random magnets. In figure 1(a), we plot the scaled structure factor $S(k,t)\langle k\rangle^2$ vs. $k/\langle k\rangle$ for the pure case and for a range of disorder amplitudes. Clearly, the functional form of the scaled structure factor appears to be unaffected by the presence of quenched disorder. In figure 1(b), we show the data of figure 1(a) on a semi-log scale, i.e., we plot $\ln(S(k,t)\langle k\rangle^2)$ vs. $k/\langle k\rangle$. It is clear that the coincidence of the scaled structure factors extends into the intermediate tail region also. The data in the extreme tail is not very reliable, in any case, as large wave-vector values are affected by the nonzero interfacial thickness. Superuniversal scaling of the structure factor has an important physical implication in that it suggests that the morphology of growing domains is unchanged inspite of trapping by disorder sites. Of course, we have only considered the lowest-order structure factor and it is not necessary that higher-order structure factors also exhibit superuniversal scaling. Nevertheless, the fact remains that the lowest-order structure factor is the experimentally relevant one and the higher-order ones are primarily of theoretical relevance. Therefore, our numerical observation has important experimental implications.

Our numerical results are not as conclusive in regards to the time-dependence of the domain length scale $R(t)$. Figure 2 plots the square of the length scale $R(t)^2$ vs. t for the pure case and for the disordered cases discussed previously. (It should be kept in mind that the length scale is measured in units of the lattice spacing.) In this figure, we plot $R(t)^2$ vs. t rather than $R(t)$ vs. t because the data for the pure case is well-known to exhibit the Lifshitz-Cahn-Allen (LCA) growth law $R(t) \sim t^{1/2}$ and this serves as a benchmark for discussing data for the disordered cases. It is evident that domain growth with quenched disorder does not exhibit the LCA growth law, except in the very early stages where domains are too small to be trapped by disorder sites. The crossover time from the LCA growth law is inversely proportional to the amplitude of disorder. Furthermore, if we plot the data of figure 2 on a log-log scale (not shown here), it is also evident that data for the disordered cases does not show a power-law behaviour at all. Unfortunately, these are the only conclusive statements we can make in this context. Inspite of extensive simulations by us [6, 12] and other authors [13], there is still no understanding of the functional form of the asymptotic growth law, or even whether there is a universal functional form valid for a wide range of disorder amplitudes. There have also been theoretical investigations of

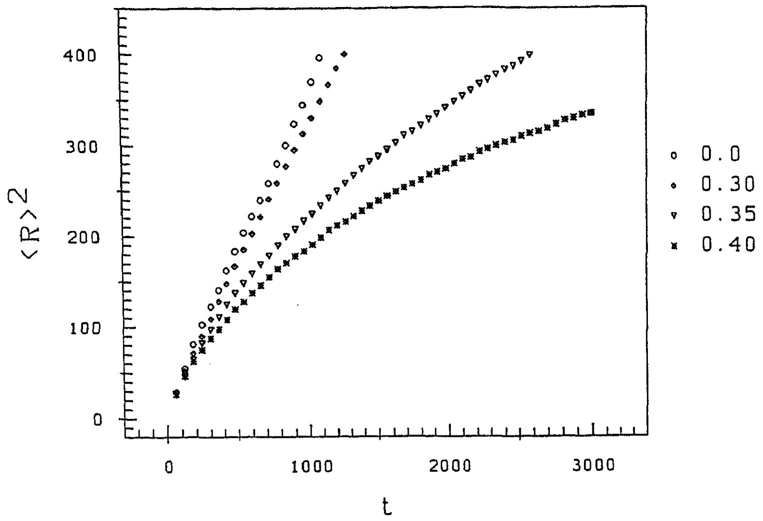


Figure 2. Square of the characteristic domain side $R(t)^2$ as a function of update time t for phase ordering dynamics in random magnets. We show data for the pure case and the disordered cases of figure 1. Symbols have the same meaning as in figure 1. (From ref. [6].)

this problem [14], which have proposed a domain growth law of the form $R(t) \sim (\ln t)^x$, where x is a universal exponent. These "logarithmic" domain growth laws are argued on the basis of thermally-activated hopping of domains over potential barriers set up by the trapping sites. However, extant numerical results do not support such domain growth laws, except in the case with quenched random fields [15].

We have also investigated the dynamics of phase separation in binary alloys with quenched disorder [16]. Our results for that case show similar trends to those discussed above for the nonconserved case, *i.e.*, the structure factor exhibits superuniversal scaling; and the asymptotic domain growth law is nonalgebraic, but we are again unable to ascertain the precise form of the universal growth law, if any.

4. Phase separation dynamics in binary fluids

We studied phase separation in binary fluids by implementing the CDS model defined by (7) on a 3-dimensional lattice of size N^3 with periodic boundary conditions. The results presented below are for the local relaxation function $f(\psi) = A \tanh \psi$ (with $A = 1.5$) and $D = 0.3$ [3]. We can associate mesh sizes with these parameter values by comparing the CDS scheme with the usual Euler discretisation scheme for partial differential equations [3]. The corresponding values are $\Delta t = 0.5$ and $\Delta x = 2.45$. In (4), we set $\bar{\eta} = 1$ and $\bar{\sigma} = 2$, following Farrell and Valls [8]. The phenomenological parameters η and σ in (7) are then fixed as $\eta = \bar{\eta}\Delta t/(\Delta x)^2 = 0.08$ and $\sigma = \bar{\sigma}\Delta t/(\Delta x)^2 = 0.17$. This should not be interpreted as a rigorous prescription for fixing parameters but rather as a rule of thumb to associate parameter values in our CDS model with those in the corresponding partial

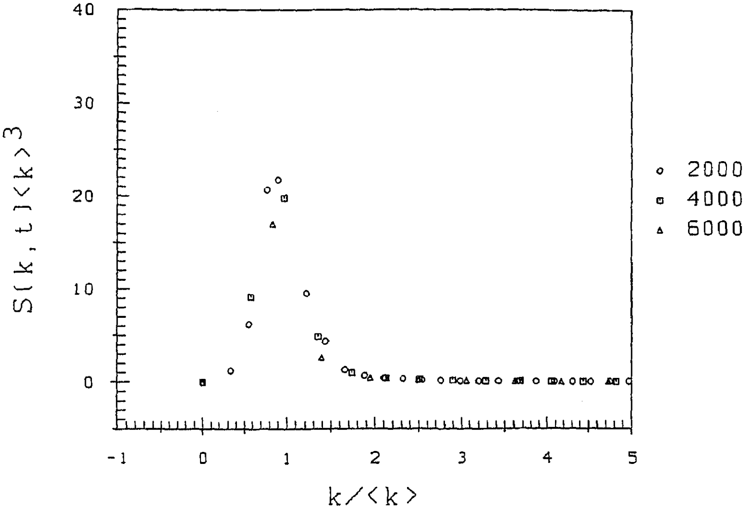


Figure 3. Superposition of data for the scaled structure factor $S(k,t)\langle k\rangle^3$ vs. $k/\langle k\rangle$ for phase separation in binary fluids. Structure factor data is obtained on 80^3 systems as an average over 40 independent runs. We superpose data from dimensionless times 2000, 4000 and 6000, denoted by the symbols indicated. (From Ref. [11].)

differential equation model. We should stress that the choice of values for viscosity parameters does not affect the asymptotic behaviour of phase separation in binary fluids and only changes the time of crossover to the asymptotic regime. Typically, higher values of viscosity suppress hydrodynamic effects and delay the onset of the asymptotic hydrodynamic regime.

Finally, we discuss the most important parameter in (7), viz., α, which fixes the strength of the coupling between the order parameter and the velocity field. We have studied domain growth as modelled by (7) for a range of values of α. In the limit $\alpha \to 0$, we recover the Lifshitz-Slyozov growth law $R(t) \sim t^{1/3}$, which characterises domain growth in Model B, as expected. In this paper, we present results for $\alpha = 0.41$, which corresponds (in the prescription defined above) to $\bar{\alpha} = 2$ in (4).

All results presented here are for the case of a critical quench, where there are equal concentrations of both components in the mixture. The structure factor is calculated for systems with $N = 80$ and as an average over 40 runs from independent initial conditions, each of which consists of the order parameter and velocity field uniformly and randomly distributed about a zero background with small amplitudes. The time-dependent characteristic length scale $R(t)$ is defined as in Eq. (9).

For purposes of comparison, we have also performed a CDS simulation for the case without hydrodynamics (i.e., Model B) on an 80^3 lattice. We use the same functional form for $f(\psi)$ and the same values for A and D as in the case with hydrodynamics.

Figure 3 superposes data for the scaled structure factor $S(k,t)\langle k\rangle^3$ vs. $k/\langle k\rangle$ from dimensionless times 2000, 4000 and 6000. The different data sets collapse neatly onto a master curve, verifying the validity of dynamical scaling in this case. We next turn our attention to the temporal behaviour of the characteristic length scale. Figure 4 plots $R(t)$

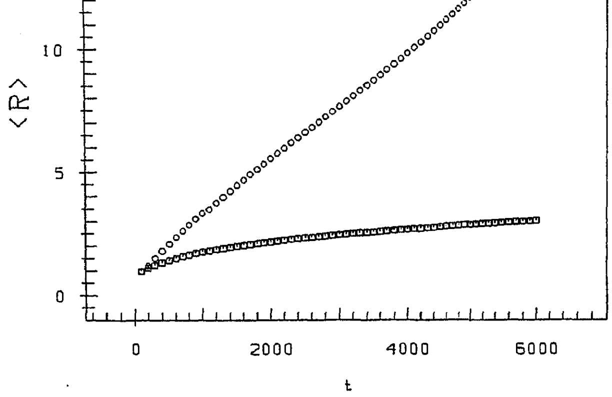


Figure 4. Characteristic domain size $R(t)$ vs. update time t for phase separation with and without hydrodynamics, denoted by circles and squares, respectively.

vs. t for our CDS model in (7) (marked by circles); and for the case without hydrodynamics (marked by squares). After an initial transient regime (which can be extended by using a smaller value of α), the domain size for the hydrodynamic case grows linearly in time, viz., $R(t) \sim t$. At times somewhat beyond those shown in figure 4, freezing sets in because of the finite size of the system. To ensure that our data is not affected by finite-size effects, we have also performed (less thorough) simulations on lattices of size 64^3 and 100^3. The results are the same as those presented here with the only difference being that the onset of freezing is delayed in larger systems. Domain growth for the case without hydrodynamics is much slower and can be easily confirmed to obey the usual Lifshitz-Slyozov growth law $R(t) \sim t^{1/3}$, though we do not show this explicitly here. Our numerical observation of temporally linear domain growth in the phase separation of binary fluids confirms a long-standing theoretical prediction by Siggia [17]. Similar results have also been obtained by Koga and Kawasaki [18] using a CDS model in which they solve for the instantaneous configuration of the fluid velocity field using the Oseen tensor approximation. There are also a number of related works, some of which we cite in Ref. [19], suggesting that there are a variety of growth regimes depending on the dimensionality and system parameters.

5. Summary and discussion

In this paper, we have attempted to acquaint the reader with some recent investigations of experimentally relevant problems in phase ordering dynamics. Initial numerical studies of idealised (i.e., pure and isotropic) phase ordering systems were greatly facilitated by Cell Dynamical System (CDS) models [3]. It is our belief that similar modelling will also help elaborate late-stage behaviour in experimentally realistic problems. To this end, we have described numerical results from CDS models of phase ordering dynamics with quenched disorder; and segregation dynamics with hydrodynamic effects.

In the context of domain growth with quenched disorder, we find that the scaled structure factor, which characterises the morphology of the coarsening system, appears to be very similar to that for the pure case. This should be experimentally verifiable and we urge experimentalists to carefully examine this aspect of the problem. Unfortunately, however, we are unable to make a conclusive statement about the asymptotic nature of the domain growth law in this problem. This is an outstanding problem which needs to be resolved but it is not at all clear that one merely needs to perform "larger" and "longer" simulations to obtain a deeper insight. As a matter of fact, we believe that some novel theoretical input is necessary to enable a better understanding of this problem.

In the context of phase separation in binary fluids, our numerical results confirm a theoretical prediction due to Siggia [17], viz., hydrodynamic modes on the surfaces of growing domains greatly enhance domain growth in the late stages, resulting in a crossover in the domain growth law from $R(t) \sim t^{1/3}$ to $R(t) \sim t$. This crossover has also been observed in a large number of experiments [1] and it is gratifying to verify it numerically.

Models similar to those described in this paper have also been used to investigate a variety of other problems in nonequilibrium statistical mechanics. Typical examples include the segregation dynamics of block copolymers [20]; population dynamics [21]; dynamics of vortices and other topological defects [22]; and the dynamics of pattern recognition [23]. In all these cases, the enhanced computational efficiency of CDS models has enabled the study of truly asymptotic behaviours. It is our belief that CDS modelling will provide us with a quantitatively accurate, and yet efficient, means of studying a host of physical phenomena.

Acknowledgements

The author would like to thank Y.Oono for a highly stimulating collaboration, which led to the initial formulation of CDS models of phase ordering dynamics. He is also deeply grateful to the individuals he collaborated with on the problems discussed in this paper.

References

[1] For reviews, see J D Gunton, M San Miguel and P S Sahni, in *Phase transitions and critical phenomena* edited by C Domb and J L Lebowitz, (Academic Press, New York, 1983) Vol. 8 p. 267
K Binder, in *Materials science and technology*
Phase transformations of materials edited by R W Cahn, P Haasen and E J Kramer (VCH, Weinheim, 1991) Vol. 5 p. 405
A J Bray, *Adv. Phys.* **43**, 357 (1994)
[2] K Binder and D Stauffer, *Phys. Rev. Lett.* **33**, 1006 (1974)
K Binder and D Stauffer, *Z. Phys.* **B24**, 407 (1976)
[3] Y Oono and S Puri, *Phys. Rev. Lett.* **58**, 836 (1987); *Phys. Rev.* **A38**, 434 (1988)
S Puri and Y Oono, *Phys. Rev.* **A38**, 1542 (1988); *J. Phys.* **A21**, L755 (1988)

S Puri, *Phys. Lett.* **A134**, 205 (1988)
A Chakrabarti and J D Gunton, *Phys. Rev.* **B37**, 3798 (1988)
[4] P C Hohenberg and B I Halperin, *Rev. Mod. Phys.* **49**, 435 (1977)
[5] S K Ma, *Modern theory of critical phenomena* (Addison-Wesley, California, 1982)
[6] S Puri, D Chowdhury and N Parekh, *J. Phys.* **A24**, L1087 (1991)
[7] J W Cahn and H E Hilliard, *J. Chem. Phys.* **28**, 258 (1958)
H E Cook, *Acta Metall.* **18**, 297 (1970)
[8] J E Farrell and O T Valls, *Phys. Rev.* **B40**, 7027 (1989)
[9] K Kaneko, *Prog. Theor. Phys.* **72**, 480 (1984)
R J Deissler, Phys. Lett. **A100**, 451 (1984)
I Waller and R Kapral, *Phys. Rev.* **A30**, 2047 (1984)
R Kapral, *Phys. Rev.* **A31**, 3868 (1985)
[10] For reviews of Coupled Map Lattice (CML) modelling, see
J P Crutchfield and K Kaneko, in *Directions in chaos*, edited by Hao Bai-Lin, (World Scientific, Singapore, 1987) pp. 272–353
K Kaneko, in *Formation, dynamics and statics of patterns*, edited by K Kawasaki (World Scientific, Singapore, 1989)
[11] S Puri and B Dunweg, *Phys. Rev.* **A45**, R6977 (1992)
[12] B Biswal, S Puri and D Chowdhury, *Physica* **A229**, 72 (1996)
[13] A J Bray and K Humayun, *J. Phys.* **A24**, L1185 (1991)
T Iwai and H Hayakawa, *J. Phys. Soc. Jpn.* **62**, 1583 (1993)
[14] D A Huse and C L Henley, *Phys. Rev. Lett.* **54**, 2708 (1985)
[15] S Puri and N Parekh, *J. Phys.* **A26**, 2089 (1993)
[16] S Puri and N Parekh, *J. Phys.* **A25**, 4127 (1993)
[17] E D Siggia, *Phy. Rev.* **A20**, 595 (1979)
[18] T Koga and K Kawasaki, *Phys. Rev.* **A44**, R817 (1991)
[19] O T Valls and J E Farrell, *Phys. Rev.* **E47**, R36 (1993)
A Shinozaki and Y Oono, *Phys. Rev.* **A45**, R2161 (1992); *Phys. Rev.* **E48**, 2622 (1993)
S Bastea and J L Lebowitz, *Phys. Rev.* **E52**, 3821 (1995)
[20] Y Oono and Y Shiwa, *Mod. Phys. Lett.* **B1**, 49 (1987)
Y Oono and M Bahiana, *Phys. Rev. Lett.* **61**, 1109 (1988)
M Bahiana and Y Oono, *Phys. Rev.* **A41**, 6763 (1990)
A Chakrabarti, R Toral and J D Gunton, *Phys. Rev.* **A44**, 6503 (1991)
[21] N Parekh and S Puri, *J. Phys.* **A23**, L1085 (1990)
[22] M Mondello and N Goldenfeld, *Phys. Rev.* **A42**, 5865 (1990); *Phys. Rev.* **E47**, 2384 (1993)
S Puri, *J. Phys.* **A24**, 3323 (1991)
H Toyoki, *Phys. Rev.* **A42**, 911 (1990)
S Puri, A J Bray and F Rojas, *Phys. Rev.* **E52**, 4699 (1995)
[23] V Banerjee and S Puri, *Int. J. Mod. Phys.* **B9**, 1159 (1995)

Noisy nonlinear coupled equations — Some new insights

J K BHATTACHARJEE, ANITA MEHTA* and J M LUCK†
Indian Association for the Cultivation of Science, Jadavpur, Calcutta 700 032, India
*Present Address: The Isaac Newton Institute for Mathematical Sciences, 20 Clarkson Road, Cambridge CB3 0HE, England
*Permanent Address: S.N. Bose National Centre for Basic Sciences, JD Block, Sector III, Salt Lake City, Calcutta 700 091, India
†C.E.A. Saclay, Service de Physique Théorique, 91191 Gif-sur-Yvette cedex, France

Abstract. We discuss the use of coupled nonlinear stochastic differential equations to model the dynamics of complex systems, and present some analytical insights into their critical behaviour. These concern in particular the role of infrared divergences which show up in a self-consistent resummation of perturbation theory (mode-coupling approximation), and their effects on critical exponents obtained in earlier work.

Keywords. Self organised criticality; Langevin equations; self consistent mode coupling.

PACS Nos 05.40; 05.70; 46.10; 64.60; 68.35

1. Introduction

The theory of self-organized criticality (SOC) [1] provided the original impetus for the study of systems with inherent complexity – systems, therefore, which could be observed in many metastable states, with, frequently, strong hysteretic behaviour. The chosen paradigm for SOC was the sandpile; it turned out, however [2, 3], that real sandpiles did not manifest the scale invariance postulated in that theory, and this proved to be the motor for a significant body of work on granular media [4].

The absence of scale invariance was postulated [5] to be due to the presence of competing dynamical mechanisms, one independent-particle and one collective, for the reorganisation of grains in a sandpile. In addition to numerous simulations [6, 7] which have supported this picture, there have been experiments recently [8] which have shown its validity.

This picture, however, necessitates the invention of new classes of models which are able to capture its intricacies. We have been involved with the formulation of discrete models which have quantified this picture, firstly, in the language of cellular automata [9] and then in that of coupled-map lattice models [10]; each of these has a representation for the two relaxation modes referred to above which turns out to be intimately related to the presence of times and lengths characteristic of the system, and hence the absence of scale invariance. At the continuum end of things, we came up with, first, a system of noisy nonlinear coupled equations at the macroscopic level [11] and next the equations [4, 12] that form the subject of the present study, which were designed to be a local representation of sandpile dynamics. The description of these equations forms the subject matter of § 2, which also contains a summary of the main results obtained thus far on their critical behaviour [12].

Section 3 concerns the analysis of these equations using a self-consistent resummation of perturbation theory (mode-coupling approximation), which provides a satisfactory explanation for the numerical results recalled in § 2. Finally, in § 4 we make some concluding remarks.

2. Coupled local equations for a driven sandpile – the MLN equations

There have been a number of recent approaches [13–17] which have attempted to model the surfaces of driven sandpiles; starting from the most basic dynamical mechanism, that of diffusion [13], these have incorporated various subtleties, ranging from lateral growth [14] to disorder [17]. However, they have nearly always involved one variable, the local height $h(\mathbf{x},t)$ of the surface, rather than any form of grain-cluster coupling. In our approach [12], however, the underlying picture is of currents of grains moving down the slope, knocking out bumps and filling in holes in clusters on the landscape. The effective coordinate representing clusters is the local height $h(\mathbf{x},t)$, while the moving grains are represented by their local density $\rho(\mathbf{x},t)$ – in a driven sandpile these coordinates will clearly be coupled, and the specific form we have chosen for this coupling follows later.

We now review some general facts about rough interfaces [18, 19]. Three critical exponents, α, β, and z, characterize the spatial and temporal scaling behaviour of a rough interface. They are conveniently defined by considering the (connected) two-point correlation function of the heights

$$S(\mathbf{x}-\mathbf{x}',t-t') = \langle h(\mathbf{x},t)h(\mathbf{x}',t')\rangle - \langle h(\mathbf{x},t)\rangle\langle h(\mathbf{x}',t')\rangle. \tag{2.1}$$

We have

$$S(\mathbf{x},0) \sim |\mathbf{x}|^{2\alpha} \quad (|\mathbf{x}| \to \infty), \qquad S(\mathbf{0},t) \sim |t|^{2\beta} \quad (|t| \to \infty), \tag{2.2}$$

and more generally

$$S(\mathbf{x},t) \approx |\mathbf{x}|^{2\alpha} F(|t|/|\mathbf{x}|^{z}) \tag{2.3}$$

in the whole long-distance scaling regime ($\mathbf{x}$ and t large). The scaling function F is universal in the usual sense; α and $z = \alpha/\beta$ are respectively referred to as the roughness exponent and the dynamical exponent of the problem.

Our stochastic dynamical equations have the following general form in one space dimension [12]

$$\partial h/\partial t = D_h \nabla^2 h - T + \eta_h(x,t), \tag{2.4}$$

$$\partial \rho/\partial t = -\nabla j + D_\rho \nabla^2 \rho + T + \eta_\rho(x,t). \tag{2.5}$$

For the time being we set

$$T = -\kappa\rho\nabla^2 h - \lambda\rho(\nabla h)_+ - \mu\rho\nabla h - \nu(\nabla h)_- \tag{2.6}$$

and

$$j = -\gamma\rho(\nabla h)_- . \tag{2.7}$$

We have introduced the notations $\nabla = \partial/\partial x$, and

$$z_+ = z\,\Theta(z) = \begin{cases} z & \text{for} \quad z \geq 0, \\ 0 & \text{for} \quad z \leq 0, \end{cases} \quad z_- = z(1-\Theta(z)) = \begin{cases} 0 & \text{for} \quad z \geq 0, \\ z & \text{for} \quad z \leq 0, \end{cases} \tag{2.8}$$

with $\Theta(z)$ being Heaviside's step function.

The content of our dynamical equations is as follows:

- The first term on the right hand side of (2.4) represents the rearrangement of clusters in the presence of an applied noise and is the conventional EW [13] term, representing diffusion.
- The second block of terms, T, represents grain-cluster exchange, where:
  - — The term $\kappa\rho\nabla^2 h$ represents diffusive motion of the heights mediated by the flowing grains.
  - — The term $\mu\rho\nabla h$ also represents intercluster motion of grains, but in a slope-dependent fashion.
  - — The term $\nu(\nabla h)_-$ represents the spontaneous generation of flowing grains whenever the local slope is larger than critical; this exists even in the absence of flowing grains and is meant to represent the effect of tilting a stationary sandpile.
  - — The term $\lambda\rho(\nabla h)_+$ is a representation of the effect of the boundary layer, in the sense that it enables the reconversion of moving to stationary grains; it compensates for the effect of the tilt term by limiting the release of flowing grains. Although originally conceived chiefly as a regulator [12], it has a definite physical meaning and, as we shall see, is responsible for the anomalous roughening of the interface.
- The first term in (2.5), $-\nabla j$, represents the variation in ρ due to the nonuniformity of the current of flowing grains, in such a way that the total number of particles is conserved. The current $j(x,t)$ has been written in (2.7) as the product of the number of mobile grains by their velocity, the latter being assumed to read $v = -\gamma(\nabla h)_-$.
- The second term in (2.5) represents the diffusive relaxation of the flowing grains, and is a crude way of representing intergrain collisions.
- The source terms $\eta_h(x,t)$ and $\eta_\rho(x,t)$ are taken to be two independent Gaussian white noises, such that

$$\begin{cases} \langle \eta_h(x,t)\eta_h(x',t')\rangle = 2\Delta_h\delta(x-x')\delta(t-t'), \\ \langle \eta_\rho(x,t)\eta_\rho(x',t')\rangle = 2\Delta_\rho\delta(x-x')\delta(t-t'). \end{cases} \tag{2.9}$$

Pouring grains onto a sandpile should be represented by noise only in ρ, while shaking it would be represented by noise in h.

We describe below the numerical results obtained in earlier work [12] on some special cases of (2.4), (2.5).

2.1 *The asymmetric situation (cases 1 to 3)*

This situation is the most commonly encountered one, of a (sloping) sandpile with a preferred direction of flow. The dynamical equations read, on setting $D_h = D_\rho = \lambda = 1$ and $\kappa = \mu = 0$,

$$\text{Cases 1 to 3}: \begin{cases} \partial h/\partial t = \nabla^2 h - T + \eta_h(x,t), \\ \partial \rho/\partial t = \nabla^2 \rho + T + \eta_\rho(x,t), \\ \quad T = -\rho(\nabla h)_+ - \nu(\nabla h)_-. \end{cases} \tag{2.10}$$

We have considered [12] the following three cases:

- Case 1: noise in h ($\Delta_h > 0$, $\Delta_\rho = 0$).
- Case 2: noise in ρ ($\Delta_h = 0$, $\Delta_\rho > 0$).
- Case 3: noise in h and ρ ($\Delta_h > 0$, $\Delta_\rho > 0$).

Non-trivial long-range spatial and temporal critical fluctuations are observed for both species h and ρ in these three cases. Case 1 is of particular interest, as it is the most directly comparable with the well-known EW [13] and KPZ [14] models. We observe a rougher behaviour of the h-profile than the two aforenamed. This effect is very pronounced in the spatial direction ($\alpha_h = 0.94 \pm 0.07$ is to be compared with $1/2$ in both cases), and still appreciable in the temporal one ($\beta_h = 0.43 \pm 0.04$ is to be compared with $1/3$ and $1/4$). This anomalous roughening is due, as we will see by the analysis in the next section, to the effect of the $\lambda\rho(\nabla h)_+$ term.

We have also investigated numerically [12] a simpler variant of the above (referred to therein as case 1a) – these were essentially identical, apart from notational differences, to the equations analysed by another group [20]. They read

$$\text{Case 1a}: \begin{cases} \partial h/\partial t = \nabla^2 h - T + \eta_h(x,t), \\ \partial \rho/\partial t = \nabla^2 \rho + T + \eta_\rho(x,t), \\ \quad T = -\mu\rho\nabla h. \end{cases} \tag{2.11}$$

The critical exponents for this case are also presented in table 1.

2.2 *The symmetric situation (cases 4 to 6)*

We now turn to the description of the symmetric version of our dynamical equations – these have the $x \leftrightarrow -x$ symmetry, and describe the surface of a sandpile which is flat on average. The dynamical equations now read in full generality

$$\text{Cases 4 to 6}: \begin{cases} \partial h/\partial t = D_h\nabla^2 h - T + \eta_h(x,t), \\ \partial \rho/\partial t = -\nabla j + D_\rho\nabla^2 \rho + T + \eta_\rho(x,t), \\ \quad T = -\kappa\rho\nabla^2 h - \lambda\rho|\nabla h| + \nu(|\nabla h| - C)_+, \\ \quad j = \gamma\rho\,\mathrm{sign}\,(\nabla h)\,(|\nabla h| - C)_+, \end{cases} \tag{2.12}$$

where C is a slope threshold. In order to investigate the critical regime, we simplify the above equations by setting $\gamma = C = 0$. We again consider three cases:

- Case 4: noise in h $(\Delta_h > 0,\ \Delta_\rho = 0)$.
- Case 5: noise in ρ $(\Delta_h = 0,\ \Delta_\rho > 0)$.
- Case 6: noise in h and in ρ $(\Delta_h > 0,\ \Delta_\rho > 0)$.

In cases 4 and 6 [12], we observe $\alpha_h \approx \beta_h \approx 0.40$ and $z_h = 1$ within error bars, just as in case 1a, whereas $\rho(x,t)$ does not exhibit divergent fluctuations: the structure factors $S_\rho(q,0)$ and $S_\rho(0,\omega)$ rather saturate to constant values for small enough values of wavevector q or frequency ω, implying the fast decay of $\rho - \rho$ correlations at long separations in space and in time. In case 5, where noise is present only in the equation for ρ, we find that h gets frozen, soon after the initial period has elapsed, into a nontrivial rough landscape, entirely inherited from the transient period, and therefore characterised by a roughness exponent $\alpha \approx 0.40$. The evolution of ρ which then takes place, with this frozen height configuration as a background, is effectively linear, and the EW exponents are observed, in spite of the background.

3. Mode-coupling analysis

In this section we provide some analytical insight into the scaling behaviour of the various cases of coupled Langevin equations discussed in the previous section. Our analysis is based on a self-consistent resummation of perturbation theory at the one-loop level, introduced long ago in the context of fluid mechanics [21], and usually referred to as the mode-coupling approximation. The main outcome will be that those models exhibit infrared singularities that control their scaling behaviour.

We illustrate this point by analysing case 1a in detail. The two coupled equations for the variables h and ρ are given in (2.11) and (2.9). We denote by $S_h(k,\omega)$ the double Fourier transform of the correlation function (2.1), such that

$$\langle h(k,\omega)h(k',\omega')\rangle = \delta(k+k')\delta(\omega+\omega')S_h(k,\omega) \tag{3.1}$$

and by $G_h(k,\omega)$ the Green's function, or propagator, of (2.11) for h, averaged over the noise, and similar notations for the ρ-field.

In the absence of coupling $(\mu = 0)$, (2.11) degenerate into two independent free (EW) models in the variables h and ρ. The Green's functions and correlation functions then read

$$\begin{aligned} G_h^{(0)-1}(k,\omega) &= -i\omega + k^2, \\ S_h^{(0)}(k,\omega) &= \frac{2\Delta_h}{\omega^2 + k^4}, \end{aligned} \tag{3.2}$$

and similar expressions for $G_\rho^{(0)}$ and $S_\rho^{(0)}$.

The effect of the coupling μ is to dress the propagator, namely to change it from the free one, given in (3.2), to the full one, $G_h(k,\omega)$. Dyson's equation yields

$$G_h^{-1}(k,\omega) = G_h^{(0)-1}(k,\omega) + \Sigma_h(k,\omega), \tag{3.3}$$

where $\Sigma_h(k,\omega)$ is the self-energy, or mass operator [22].

Within the self-consistent one-loop (mode-coupling) approximation, where the internal lines of the self-energy graph are fully dressed, we have

$$\Sigma_h(k,\omega) = \mu^2 \int \frac{dk'}{2\pi} \int \frac{d\omega'}{2\pi} S_\rho(k',\omega') k(k-k') G_h(k-k',\omega-\omega'). \tag{3.4}$$

Now we make the following scaling ansatz about $\Sigma_h(k,\omega)$ and $S_h(k,\omega)$

$$\begin{aligned} \Sigma_h(k,\omega) &\approx k^{z_h}\sigma(\omega/k^{z_h}), \\ S_h(k,\omega) &\approx k^{-1-2\alpha_h-z_h} S(\omega/k^{z_h}), \end{aligned} \qquad (k>0) \tag{3.5}$$

and similarly for $\Sigma_\rho(k,\omega)$ and $S_\rho(k,\omega)$, with exponents z_ρ and α_ρ. The determination of z_h involves the computation of the static (zero-frequency) self-energy $\Sigma_h(k) = \Sigma_h(k,\omega=0)$. We assume $z_h < 2$, so that $\Sigma_h(k)$ dominates k^2 in the scaling region [this hypothesis will be checked a posteriori]. Hence in (3.3) we have

$$G_h(k,\omega)^{-1} \approx -i\omega + \Sigma_h(k,\omega). \tag{3.6}$$

Accordingly, we explore the integral on the right-hand-side of (3.4) at zero external frequency. The integral has an infrared divergence when the momentum k' of the correlation function goes to zero. Indeed in this limit $k(k-k') \approx k^2$, and

$$\begin{aligned} G_h^{-1}(k-k',-\omega') &\approx i\omega' + \Sigma_h(k-k') \\ &\approx i\omega' + \Sigma_h(k) \\ &\approx \Sigma_h(k). \end{aligned} \tag{3.7}$$

Because of the scaling laws (3.5) we effectively explore the low-frequency behaviour when we explore low values of k'. The self-energy of (3.4) now becomes

$$\begin{aligned} \Sigma_h(k) &\approx \frac{\mu^2 k^2}{\Sigma_h(k)} \int \frac{dk'}{2\pi} \int \frac{d\omega'}{2\pi} S_\rho(k',\omega') \\ &\sim \frac{\mu^2 k^2}{\Sigma_h(k)} \int \frac{dk'}{2\pi} |k'|^{-1-2\alpha_\rho}, \end{aligned} \tag{3.8}$$

since

$$\int \frac{d\omega}{2\pi} S_h(k,\omega) \sim |k|^{-1-2\alpha_h}, \tag{3.9}$$

and a similar formula for the ρ-field, as a consequence of (3.5). The integral on the right-hand-side of (3.8) clearly exhibits an infrared divergence. For $\alpha_\rho \geq 0$, the integral diverges at small k'. If we regularise the integral by introducing as an infrared cutoff a smallest wavevector k_0, (3.8) yields

$$\Sigma_h(k) \sim \mu |k| k_0^{-\alpha_\rho}. \tag{3.10}$$

The above result at zero frequency implies $z_h = 1$; an identical analysis yields $z_\rho = 1$.

Now, in order to determine α_h, we investigate the correlation function $S_h(k,\omega)$. Again within the self-consistent one-loop approximation, we have

$$\begin{aligned} S_h(k,\omega) &\approx \frac{1}{\omega^2+k^4} + \mu^2 \int \frac{\mathrm{d}k'}{2\pi} \int \frac{\mathrm{d}\omega'}{2\pi} S_\rho(k',\omega')(k-k')^2 S_h(k-k',\omega-\omega') \\ &\approx \frac{1}{\omega^2+k^4}\left[1+\mu^2 \int \frac{\mathrm{d}k'}{2\pi} \frac{(k-k')^2}{|k-k'|^{1+2\alpha_h}} \frac{1}{|k'|^{1+2\alpha_\rho}} \frac{\Sigma_\rho(k')+\Sigma_h(k-k')}{\omega^2+(\Sigma_\rho(k')+\Sigma_h(k-k'))^2}\right]. \end{aligned} \tag{3.11}$$

Using (3.9), we have

$$\frac{1}{k^{1+2\alpha_h}} \approx \frac{1}{k^2} + \mu^2 \int \frac{\mathrm{d}k'}{2\pi} \frac{(k-k')^2}{(k-k')^{1+2\alpha_h}} \frac{1}{k'^{1+2\alpha_\rho}}. \tag{3.12}$$

The low-momentum divergence of the integral on the right-hand-side of (3.12) implies that the integral scales as $|k|^{1-2\alpha_h} k_0^{-2\alpha_\rho}$, which is much smaller than $k^{-1-2\alpha_h}$ in the long-wavelength limit. Thus α_h remains at its EW value, namely $\alpha_h = 1/2$. An identical argument on the ρ-correlation yields $\alpha_h = \alpha_\rho = 1/2$ and $z = 1$ in case 1a.

A variant of case 1a is the model studied by Bouchaud *et al* [20]. In our notation, this model reads

$$\begin{cases} \partial h/\partial t = \nabla^2 h + c\nabla h + \mu\rho\nabla h + \eta_h(x,t), \\ \partial \rho/\partial t = \nabla^2 \rho - c\nabla h - \mu\rho\nabla h. \end{cases} \tag{3.13}$$

The zeroth-order (free) Green's function and correlation function for h are respectively

$$\begin{aligned} G_h^{(0)}(k,\omega) &= \frac{1}{-i\omega+k^2+ick}, \\ S_h^{(0)}(k,\omega) &= \frac{2\Delta_h}{(\omega-ck)^2+k^4}. \end{aligned} \tag{3.14}$$

The coupling between both fields can be viewed as a coloured noise acting on ρ. The zeroth-order correlation function for ρ is

$$S_\rho^{(0)}(k,\omega) = \frac{2c^2\Delta_h k^2}{\left((\omega-ck)^2+k^4\right)(\omega^2+k^4)}, \tag{3.15}$$

hence

$$\int S_\rho^{(0)}(k,\omega)\frac{\mathrm{d}\omega}{2\pi} = \frac{2c^2\Delta_h}{k^2(4k^2+c^2)}. \tag{3.16}$$

The full Green's function is then again given by (3.3), with the self-energy as in (3.4). The low-momentum region produces a divergent integral, because of (3.9), so that the conclusion $z_h = 1$ is still valid. The analysis of the correlation function follows in a manner similar to that done for case 1a, and leads to $\alpha_h = 1/2$.

Bouchaud *et al* [20] applied standard perturbation theory to their equations (3.13). They did not meet the infrared singularity described above, since this phenomenon can only appear through treatments which go beyond strict perturbation theory, such as our self-consistent resummation procedure. Hence the conclusion that the model (3.13) is free (EW-like) on a perturbative basis is incorrect.

To close up this analysis, we come back to our case 1. We observe that the transfer term

$$T = -\nu(\nabla h)_- - \lambda\rho(\nabla h)_+ \tag{3.17}$$

can be thought of as a formal infinite series by invoking a suitable representation for Heaviside's Θ-function. We are thus led to consider the following more general structure for the transfer term

$$T = -\nu\nabla h - \lambda\rho\nabla h - \sum_{n=1}^{\infty}\nu_n(\nabla h)^{n+1} - \rho\sum_{n=1}^{\infty}\lambda_n(\nabla h)^{n+1}. \tag{3.18}$$

Dimensional analysis along the lines of ref. [12] shows that, among the terms in the sums, only $\nu_1(\nabla h)^2$ is relevant in one space dimension. Now consider the calculation of $\Sigma_h(k)$, within the above self-consistent one-loop approach. Two contributions are to be taken into account. We have already dealt with one of them in (3.4). That forces $z_h = 1$, independently of α_h. The other contribution is of the KPZ variety [14], for which a symmetry property, referred to as Galilean invariance, requires $\alpha_h + z_h = 2$ [18,19]. This in turn forces $\alpha_h = 1$ and $\beta_h = 1$.

The numerical work [12] for case 1 suggests $\alpha_h \approx 1$ and $\beta_h \approx 1/2$, consistent with $z_h \approx 2$. One possible, although rather exotic, explanation for the difference between the analytical prediction $z_h = 1$ and the numerical estimate $z_h \approx 2$ is the phenomenon observed recently [23] in a suitably regularized mean-field version of the KPZ problem. These authors have shown that the dynamics of the approach to the stationary state is much slower than the linear-response dynamics in the stationary state.

Finally, for our cases 4 and 6, which bear a close resemblance to the models studied by Krug and Spohn [19], the present work yields $\alpha_\rho = 1/2$ and $z_h = 1$, as was indeed found numerically in [12]. The situation of case 5 was already discussed in § 2.

4. Discussion

In the above, we have presented analytical insights on the behaviour of a system of coupled noisy nonlinear equations put forward earlier [4, 12] as a model of sandpile dynamics. The anomalous roughening seen in case 1, in particular, is seen to be due predominantly to the effect of the term $\lambda\rho(\nabla h)_+$; this has a simple physical explanation in the context of the observed agreement of our results with the experimental work of Kurnaz *et al* [24]. The authors of that work found, for the case of a rotated and shaken sandpile, exponents $\alpha_h = 0.92 \pm 0.05$ and $\beta_h = 0.48 \pm 0.16$ (in close agreement with case 1 in our table), and attributed this to the effects of interstitial fluids in their system. The effect of the $\lambda\rho(\nabla h)_+$ term is precisely to put back "roughness" on the profile of the pile at a rate proportional to the density ρ of mobile grains. We see therefore that our analysis above, among other things, provides a rather appealing picture of an experimental scenario.

Table 1. Critical exponents α, β, and z for both fields $h(x,t)$ and $\rho(x,t)$, measured from numerical simulations of the coupled stochastic equations in the various cases described in text (after ref. [12]). The exact EW and KPZ values are recalled for comparison. A hyphen means that the corresponding quantity is in our opinion not critical. An asterisk means that the exponent z cannot be accurately evaluated from the available data on α and β.

Model	Species	α	β	z
EW	h	1/2	1/4	2
KPZ	h	1/2	1/3	3/2
(1) asymmetric noise in h	h	0.94 ± 0.07	0.43 ± 0.04	2.2 ± 0.4
	μ	0.22 ± 0.08	0.07 ± 0.02	*
(1a) no-tilt noise in hand in ρ	h	0.36 ± 0.02	0.41 ± 0.08	0.9 ± 0.2
	ρ	0.39 ± 0.05	0.27 ± 0.04	1.4 ± 0.4

Acknowledgements

One of the authors (AM) acknowledges the hospitality of the Isaac Newton Institute for Mathematical Sciences, where part of this work was carried out.

References

[1] P Bak, C Tang and K Wiesenfeld, *Phys. Rev. Lett.* **59**, 381 (1987); *Phys. Rev.* **A38**, 364 (1988)
[2] Anita Mehta and G C Barker, *Rep. Prog. Phys.* **57**, 383 (1994)
[3] H M Jaeger and S R Nagel, *Science* **255**, 1523 (1992)
[4] *Granular Matter: An interdisciplinary approach* edited by Anita Mehta (Springer Verlag, New York, 1993)
[5] Anita Mehta, *The physics of powders*, in *Correlations and Connectivity* edited by H E Stanley and N Ostrowsky (Kluwer Academic Press, Dordrecht, 1990)
[6] Anita Mehta and G C Barker, *Phys. Rev. Lett.* **67**, 394 (1991)
[7] G C Barker and Anita Mehta, *Phys. Rev.* **A45**, 3435 (1992)
[8] V Frette, K Christensen, A Malthe-Sørenssen, J Feder, T Jøssang and P Meakin, *Nature (London)* **379**, 49 (1996)
[9] Anita Mehta and G C Barker, *Europhys. Lett.* **27**, 501 (1994)
[10] G C Barker and Anita Mehta, *Phys. Rev.* **E53**, 5704 (1996)
[11] Anita Mehta, R J Needs and S Dattagupta, *J. Stat. Phys.* **68**, 1131 (1992)
[12] Anita Mehta, J M Luck and R J Needs, *Phys. Rev.* **E53**, 92 (1996)
Anita Mehta, G C Barker, J M Luck and R J Needs, *Physica* **A224**, 48 (1996)
[13] S F Edwards and D R Wilkinson, *Proc. R. Soc. London* **A381**, 17 (1982)
[14] M Kardar, G Parisi and Y Zhang, *Phys. Rev. Lett.* **56**, 889 (1986)
[15] T Hwa and M Kardar, *Phys. Rev. Lett.* **62**, 1813 (1989)
[16] G Grinstein and D H Lee, *Phys. Rev. Lett.* **66**, 177 (1991)
[17] J Toner, *Phys. Rev. Lett.* **66**, 679 (1991)
[18] T Halpin-Healy and Y C Zhang, *Phys. Rep.* **254**, 215 (1995)
A L Barábasi and H E Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, 1995)
[19] J Krug and H Spohn, in *Solids far from equilibrium* edited by C Godřeche (Cambridge University Press, 1992)

[20] J-P Bouchaud, M E Cates, J Ravi Prakash and S F Edwards, *J. Phys.* **4**, 1383 (1994); *Phys. Rev. Lett.* **74**, 1982 (1995)
[21] R Kraichnan, *J. Fluid Mech.* **5**, 497 (1959)
[22] G D Mahan, *Many-particle physics* (Plenum, New York, 1981)
J P Blaizot and G Ripka, *Quantum theory of finite systems* (M.I.T. Press, 1986)
[23] M Marsili and A J Bray, *Phys. Rev. Lett.* **76**, 2750 (1996)
[24] M L Kurnaz, K V McCloud and J V Maher, *Fractals* **1**, 1008 (1993)
M L Kurnaz and J V Maher, *Phys. Rev.* **E53**, 978 (1996)

—journal of
physics

February 1997
pp. 759–773

Nonlinear fluctuating hydrodynamics and sequence of time scales of relaxation in supercooled liquids

SHANKAR P DAS
School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Abstract. Nonlinear fluctuating hydrodynamic (NFH) models for relaxation in the supercooled liquid are considered. Recent results on self consistent mode coupling theory for the slow relaxation of density fluctuations are analyzed to explain the glassy dynamics. The relaxation mechanisms for different types of models with and without wave vector dependences are discussed. For the schematic models where all wave vector dependences are dropped a sequence of time scales enters the relaxation process. For the non-ergodicity parameter very close to the ideal transition point is scaled by an exponent equal to 1/2. This is demonstrated here through an analysis of the mode-coupling equations for the wave vector dependent models that follow from equations of NFH.

Keywords. Glass transition; ergodicity; correlation function.

PACS Nos **64.70; 05.60; 64.60**

1. Introduction

In the last ten years a good amount of progress has been achieved on the understanding of the dynamics of supercooled liquids using models derived from the nonlinear fluctuating hydrodynamic equations for the compressible fluid [1, 2]. In these models the transport coefficients for the supercooled liquid are obtained in terms of self-consistent expressions of the hydrodynamic correlation functions. This constitute a feedback mechanism [3, 4] giving rise to slow relaxation in the supercooled liquid. In the self consistent mode coupling models for glass transition, the feed back effects from the terms involving the slowly decaying density fluctuations are analyzed assuming that they produce the dominant contribution at supercooled states. In this paper we will focus our discussion on the nonlinear hydrodynamic models, specially with structural effects taken into account.

The present form of the mode-coupling theory (MCT) for supercooled liquid dynamics was initiated from the work of Leutheusser by showing that a model obtained from the kinetic theory of dense fluids exhibits an ergodic-nonergodic transition which has features similar to a liquid-glass transition. The model leads to a viscosity which diverges as $(T - T_0)^{-\alpha}$ as the temperature T approaches the ideal glass transition T_0. The exponent α was found to take values $\alpha \sim 2$. Later it was demonstrated [5] that the sharp transition is cutoff due to a mechanism arising from coupling of density and current fluctuations in a compressible fluid and this keeps the system ergodic at all temperatures. However the analysis done in ref. [5] also showed that the feed back mechanism from the coupling of density fluctuations does cause a substantial enhancement of the viscosity. Thus although the sharp transition is cutoff there are some strong remnants of it with a qualitative change in the supercooled liquid dynamics around a temperature T_0 higher than the usual glass transition temperature. Indeed such behavior was observed in a number of liquids [6, 7] termed as the fragile liquids. In a number of subsequent works [7] the cutoff mechanism responsible for ergodic behavior at the longest time scale in the supercooled liquid was simply ignored and the relaxation behavior around the point which involves a sharp transition from ergodic to non ergodic behavior was analyzed. The simplest example of this type is the Leutheusser model which consider a schematic form of the mode-coupling theory with all q dependence being dropped, but only the main characteristics of the dynamic instability being preserved. A sequence of time scales enter the relaxation process in the schematic model. However, when the full wave vector dependence in the models are taken into account , through a realistic static structure factor this aspect changes somewhat. We discuss these in the present work by presenting a simple analysis of the wave vector dependent model here.

We will divide the discussion mainly into two parts namely the idealized models with a sharp transition and the fully self-consistent model in which there is no sharp transition in the dynamic behavior of the liquid. In § 2 we first discuss the nonlinear fluctuating hydrodynamic equations for a compressible fluid with a simple choice of conserved densities as slow variables and discuss how the mode coupling model is obtained from these equations. In § 3, we discuss the idealized model which has a dynamic instability giving rise to the so called glass transition and discuss the various time scales that enter the relaxation process. In the next section we discuss the fully self consistent model with the cutoff mechanism for restoring ergodicity and discuss some recent results regarding the nature of the kernel for the cutoff mechanism. We end the paper with a small discussion.

2. Fluctuating hydrodynamics of compressible liquid

In the present work we will consider the set of hydrodynamic variables [5] as the mass density ρ, the momentum density $\mathbf{g}$, and the flow velocity $\mathbf{v}$ defined through the nonlinear constraint

$$\mathbf{g} = \rho \mathbf{v}. \tag{1}$$

The equations of motion for the hydrodynamic variables, are obtained using the well known Zwanzig–Mori [8] formalism and are valid for small and finite wavelengths. The equation for ρ is given by

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{g} \tag{2}$$

and that for $\mathbf{g}$ is the generalized nonlinear Navier Stokes equation with thermal noise

$$\frac{\partial g_i}{\partial t} = -\rho \nabla_i \frac{\delta F_u}{\delta \rho} - \sum_j \nabla_j (g_i v_j) - \sum_j \int \mathrm{d}x' L_{ij}(x - x') \frac{\delta F}{\delta g_j(x')} + \Theta_i, \tag{3}$$

where $F_u[\rho(x)]$ is the potential energy part of the effective Hamiltonian F defined [9] as

$$F = \frac{1}{2} \int \mathrm{d}^3 x g^2(x)/\rho(x) + F_u. \tag{4}$$

Following the usual forms common in the density functional theories, F_u is taken as an expansion of an inhomogeneous equilibrium liquid

$$(\beta m)F_u[\rho(x)] = \int \mathrm{d}x\rho(x)\{\ln[\rho(x)/\rho_0] - 1\}, +F_{\mathrm{int}}[\rho], \tag{5}$$

where the first term is the ideal gas entropy term and the interaction term F_{int} to lowest order can be obtained (up to a constant) as

$$F_{\mathrm{int}}[\rho] = -\frac{1}{2m}\int \mathrm{d}^3x\mathrm{d}^3x' c^{(2)}(x-x')\delta\rho(x)\delta\rho(x') \tag{6}$$

with $\delta\rho(x,t) = \rho(x,t) - \rho_0$ and $\beta = 1/k_BT \cdot c(x)$ is the equilibrium two particle correlation function for the liquid. In an isotropic fluid the bare transport matrix $L_{ij}(\mathrm{x})$ is related to the Gaussian noise Θ_i through the fluctuation dissipation relation

$$\langle\Theta_i(\mathrm{x},t)\Theta_j(\mathrm{x}',t')\rangle = 2k_BTL_{ij}(\mathrm{x})\delta(\mathrm{x}-\mathrm{x}')\delta(t-t'). \tag{7}$$

For an isotropic fluid the bare transport matrix $L_{ij}(\mathrm{x})$ or its Fourier transform

$$L_{ij}(q) = \int \mathrm{d}x\, \mathrm{e}^{i\mathrm{q}\cdot\mathrm{x}}L_{ij}(x) \tag{8}$$

can be expressed in terms of two independent transport coefficients given by

$$L_{ij}(q) = q_iq_j\Gamma^0(q) + [q^2\delta_{ij} - q_iq_j]\eta^0(q). \tag{9}$$

In the small wavenumber limit, $\Gamma^0(q)$ and $\eta^0(q)$ are the bare longitudinal and shear viscosities respectively. Since we will be applying these equations for finite wavelengths here, more generalized expressions for these quantities are obtained with the use of Enskog type models [10]. These reflect the short time properties referring to uncorrelated random collisions in the system.

In order to investigate the effects the nonlinearities in the hydrodynamic equations will have on the transport properties of the fluid, a field theory of the Martin–Siggia–Rose (MSR) [11] type was used in ref. [5]. The advantage of using the MSR field theory here is that the renormalized expressions for the various quantities are obtained in a self consistent manner in terms of the full correlation functions and is very useful in demonstrating the feedback mechanism that results in slow relaxation at supercooled densities. The field theoretic treatment of the problem is done following the techniques developed in the theories of dynamical critical point phenomena and for this the interested reader is referred to the papers cited under ref. [11]. For the analysis as applied to the theory of the compressible fluids see ref. [5, 12]. In the present discussion we give a very brief account of the steps used in obtaining the mode coupling models as applied for supercooled liquids. The fully renormalized theory of the hydrodynamic correlation functions are obtained in terms of the self energy matrix Σ defined through the Dyson equation

$$G^{-1}(\mathbf{q},\omega) = G_0{}^{-1}(\mathbf{q},\omega) - \Sigma(\mathbf{q},\omega), \tag{10}$$

where G_0 refers to the matrix of correlation functions obtained from the equations of linearized hydrodynamics. The main quantity of interest here is the density auto

$$G_{\rho\rho}(q,t) = \int \mathrm{d}(\mathbf{x}-\mathbf{x}')\mathrm{e}^{i\mathbf{q}\cdot(\mathbf{x}-\mathbf{x}')}\langle\delta\rho(\mathbf{x},t)\delta\rho(\mathbf{x}',0)\rangle, \tag{11}$$

where the angular bracket refers to the average over the stationary states. In ref. [5] the following form for the Fourier–Laplace transform of $G_{\rho\rho}(\mathbf{x},t)$ normalized with respect to its equal time value is obtained in the small q and ω limit

$$\phi(q,z) = \frac{z + iD_L{}^R(q,z)}{z^2 - \Omega_q^2 + iD_L{}^R(q,z)[z + i\gamma(q,z)]}. \tag{12}$$

Here $\Omega_q^2 = q^2[\beta m S(q)]^{-1}$ and $D_L{}^R(q,z)$ is the renormalized longitudinal viscosity. The Laplace transform of $\phi(t)$ is defined as

$$\phi(z) = (-i)\int_0^\infty \mathrm{d}t\, \mathrm{e}^{izt}\phi(t), \quad \mathrm{Im}(z) > 0. \tag{13}$$

Similarly the Laplace transform for the transverse current fluctuation (normalized with respect to its equal time value) is given by

$$\phi^T(q,z) = \frac{1}{z + i\eta^R(q,z)}, \tag{14}$$

where $\eta^R(q,z)$ is the renormalized shear viscosity. In the formulation of the MSR type field theory the renormalized memory kernels on the R.H.S of (12) and (14) have the mode coupling contributions at the one loop level respectively given by

$$\begin{aligned} D_L{}^{mc}(q,t) =& \lambda_0 \int \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}[\{\hat{q}\cdot\mathbf{k}\}c(k) + \{\hat{q}\cdot(\mathbf{q}-\mathbf{k})\}c(|\mathbf{q}-\mathbf{k}|)]^2 \\ & \times G_{\rho\rho}(\mathbf{q}-\mathbf{k},t)G_{\rho\rho}(\mathbf{k},t) \end{aligned} \tag{15}$$

and

$$\eta^{mc}(q,t) = \lambda_0 \int \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}[c(k) - c(\mathbf{q}-\mathbf{k})]^2 k^2(1-u^2)G_{\rho\rho}(\mathbf{q}-\mathbf{k},t)G_{\rho\rho}(\mathbf{k},t), \tag{16}$$

where $\lambda_0 = (2\beta m^4 \rho_0)^{-1}$ and $u = \hat{q}\cdot\hat{k}$ while $\hat{q}$ is the unit vector along the direction of $\mathbf{q}$. The quantity $\gamma(q,z)$ in the R.H.S of (12) arises from the coupling between the density fluctuations and current fluctuations in a compressible fluid. In the asymptotic limit when the viscosity become large due to the feedback coming from mode coupling contributions, the density autocorrelation function given by (12) develops a pole at

$$z + i\gamma(q,z) = 0. \tag{17}$$

Thus ergodicity is restored over a time scale of $1/\gamma(0,0)$, representing the longest time scale in the relaxation of the density correlation function. However, in the simple model where all the nonhydrodynamic corrections coming from time scales of the order of q^2

are ignored the simplified model has a dynamic instability similar to the liquid-glass transition. In the next section we will consider this model first to demonstrate the several relaxation processes that can be included in the context of mode coupling theory and the consequences of a proper wave vector dependence in the theory.

3. Effects of the feedback mechanism and sequence of time scales of relaxation

3.1 *Schematic models without wavevector dependence*

In this section we consider the simple model with the sharp instability showing a transition to an ideal glassy phase. In this regard several different types of models have been investigated. First we consider the schematic models where all the wave vector dependences in the mode coupling kernels for transport coefficients are dropped and the density autocorrelation function $\phi(z)$ following (12) is given by the functional equation

$$\frac{-\Omega_0^2\phi(z)}{1 - z\phi(z)} = z + iD_L(z), \tag{18}$$

where Ω_0 is a microscopic frequency. In the schematic models the renormalized expression for the viscosity $D_L(z)$ is obtained as

$$D_L(z) = D_L^0 + \Omega_0^2 \int_0^{\infty} \mathrm{d}t\, \mathrm{e}^{izt} H[\phi(t)], \tag{19}$$

where D_L^0 is the bare viscosity governing the microscopic dynamics. In studying the feedback mechanism for slow relaxation in the supercooled liquid the memory kernel H is expressed as a functional of the density correlation function $\phi(t)$ [7, 13] given by

$$H[\phi(t)] = \sum_{n=1}^{N} c_n \phi^n(t). \tag{20}$$

In the model considered by Leutheusser $H[\phi] = c_2\phi^2$. The more general model with $c_1 = 0$ was first introduced in ref. [1]. The importance of the linear term was first pointed out by Goetze [14]. In the above expressions for the memory kernel all wave vector dependences are assumed to be weak and are suppressed. The basic assumption in the analysis of the equation (18) is that depending on the the kernel $H[\phi]$ i.e. the coefficients c_ns there is a time range over which $\phi(t)$ is approximately time independent following the form

$$\phi(t) = f + (1 - f)\phi_\nu(t), \tag{21}$$

where f is the value of $\phi(t)$ in some metastable state which is yet to be specified. For time scales where the inequality $|z\phi_\nu(z)| \ll 1$ is valid, we obtain to leading order from equation (18)

$$\frac{\Delta_0}{z} + \Delta_1\phi_\nu(z) - (1 - f)H'(f)z{\phi_\nu}^2(z) + \frac{(1 - f)^2}{2}H''(f)L_z[{\phi_\nu}^2(t)] = 0 \tag{22}$$

$$\Delta_1 = \frac{\partial}{\partial f}[(1-f)H(f) - f]. \tag{23}$$

We have used the notation $H''(f) = (\partial^2 H(f))/\partial f^2$ etc. and L_z stands for the Laplace transform of the function in the argument as defined in equation (13) with variable z. An ideal metastable state is obtained when both Δ_0 and Δ_1 are zero giving a solution for the decaying function $\phi_\nu(t)$. We can determine f by setting $\Delta_1 = 0$. This is obtained for

$$H(f) + 1 = \frac{C}{(1-f)}, \tag{24}$$

with C is a constant not depending on f. At this condition $\Delta_0 = (C-1)/(1-f)$. An ideal state is obtained when for a given set of c_n we have $\Delta_0(c_n) = 0$. Thus the ideal transition takes place with for $C = 1$ giving $\Delta_0 = 0$. For the Leutheusser model we obtain that $f = 1/2$ and c_2 has a critical value of $c_2^* = 4$. For the model by Goetze [14] we have linear term in the kernel, i.e.

$$H[\phi] = c_1\phi + c_2\phi^2, \tag{25}$$

with a line of critical values satisfying

$$c_1^* = 2\sqrt{c_2^*} - c_2^*. \tag{26}$$

For higher order models there are critical surfaces. Note that with $\Delta_1 = 0$ the equation (22) can be written as

$$\frac{\Delta_0}{z} - \frac{C}{(1-f)} z\phi_\nu^2(z) + \frac{(1-f)^2}{2} H''(f) L_z[\phi_\nu^2(t)] = 0. \tag{27}$$

For the high frequency region where the term proportional to Δ_0 can be neglected i.e.

$$(|\Delta_0|)^{1/2} \ll |z\phi_\nu(z)| \tag{28}$$

the above equation reduces to the simple form,

$$z\phi_\nu^2(z) - \lambda L[\phi_\nu^2(t)] = 0 \tag{29}$$

with $\lambda = ((1-f)^3/2C)\,H''(f)$. This equation is satisfied by the function

$$\phi_\nu(t) = A(t\Omega_0)^{-a}. \tag{30}$$

The exponent a is given in terms of the parameter λ by the equation

$$\frac{\Gamma^2(1-a)}{\Gamma(1-2a)} = \lambda, \tag{31}$$

where Γ is the gamma function. The inequality (28) and the solution (30) obviously holds in the time region

$$\Omega_0{}^{-1} \ll t \ll \tau_a \equiv |\Delta_0|^{-1/2a}. \tag{32}$$

The time scale τ_a diverges at the ideal transition point i.e. $\Delta_0 = 0$. Therefore $\phi(t)$ decays algebraically towards the metastable value f. For $\Delta_0 \neq 0$ i.e. points away from the transition the dynamics is quite different for the cases $\Delta_0 > 0$ and $\Delta_0 < 0$. For $\Delta_0 > 0$, equation (27) has the solution $\phi_\nu(z) \sim \Delta_0^{1/2}/z$ as $z \to 0$ and

$$f = f_0 + c(\Delta_0)^{1/2} + O(\Delta_0), \tag{33}$$

where f_0 is the value of f at the ideal transition point and c is a constant. For $\Delta_0 < 0$ we have from equation (27) a solution for $\phi_\nu(z)$ which is more singular that $1/z$ as $z \to 0$ and has the following form in the time regime

$$\phi_\nu(t) = -B(t/\tau_\alpha)^b, \tag{34}$$

where B is a positive constant. Similar to a in (30), b satisfies the equation

$$\frac{\Gamma^2(1+b)}{\Gamma(1+2b)} = \lambda. \tag{35}$$

Equation (34) is referred to as the von-Schweidler relaxation law. However this relaxation eventually violates the inequality $|z\phi_\nu(z)| \ll 1$ and this decay mechanism is valid in the region

$$\tau_a \ll t \ll \tau_\alpha, \tag{36}$$

where τ_α is given by

$$\tau_\alpha = |\Delta_0|^{-(1/2a+1/2b)}. \tag{37}$$

For $t \geq \tau_\alpha$, the system enters into the α relaxation regime. In these regime a fully analytic solution of the mode coupling equations is still not available. Numerical solution for a set of parameters [13] has shown that the solution is well fit by a stretched exponential

$$\phi(t) = f\,\mathrm{e}^{-(t/\tau_\alpha)^\beta}. \tag{38}$$

The final decay of the correlation function takes place over the time scale, given by $1/\gamma$. We show in figure 1 the sequence of times scales that enter the relaxation process in the mode coupling models that uses the schematic forms for the memory kernel in terms of the density correlation function.

3.2 *Effects of including proper structure factor*

In this case we consider the set of equations (for different q values),

$$\frac{-\Omega_q^2\phi(q,z)}{1 - z\phi(q,z)} = z + iD_L(q,z), \tag{39}$$

where the memory kernel representing the transport behavior is given by

$$D_L(q,z) = D_L^0(q) + \Omega_0^2 \int_0^\infty \mathrm{d}t\,\mathrm{e}^{izt} H_q[\phi(q,t)], \tag{40}$$

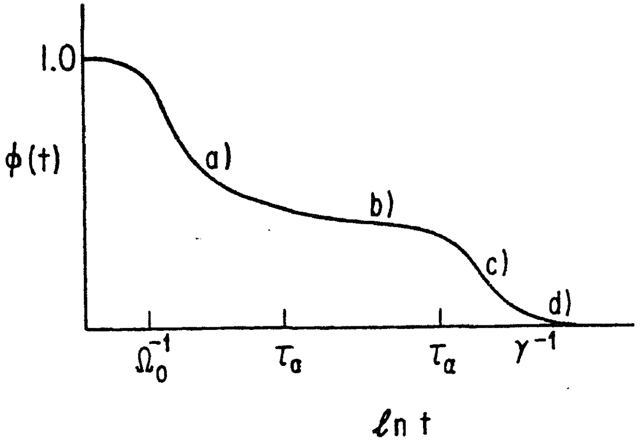


Figure 1. Schematic plot of the sequence of time scales in the relaxation predicted by MCT; (**a**) power-law decay $t^{-\alpha}$, (**b**) von Schweidler relaxation $-Bt^b$, (**c**) primary relaxation $e^{-(t/\tau)^\beta}$, (**d**) exponential relaxation $e^{-\gamma t}$.

where $D_L^0(q)$ is the bare viscosity [15] governing the microscopic dynamics over length scale corresponding to wave vector q. The function $H_q[\phi]$ now depends on all the values of ϕ at different q values. In general we can have the kernel in the form

$$H_q[\phi(t)] = \sum_{k_1} V^{(1)}(q,k_1)\phi(k_1,t) + \sum_{k_1,k_2} V^{(2)}(q,k_1,k_2)\phi(k_1,t)\phi(k_2,t) + \sum_{k_1,k_2,k_3} V^{(3)}(q,k_1,k_2,k_3)\phi(k_1,t)\phi(k_2,t)\phi(k_3,t) + \cdots . \tag{41}$$

In the model obtained using the nonlinear fluctuating hydrodynamics described in equation (15) $V^{(1)}$ happens to be zero. Subsequently, Kim [16] has considered models of fluctuating hydrodynamics that obtains a term in the kernel linearly proportional to the density correlation function. We look for a solution of the above equations in the form

$$\phi(q,t) = f_q + (1-f_q)\phi_\nu(q,t). \tag{42}$$

If f_q happens to be a nonzero set of values beyond a certain density then this corresponds to an ideal non ergodic state where the density correlation function decays to a finite value instead of decaying to zero. Following an analysis similar to what is described in the previous section for the schematic model, we obtain an equation

$$\frac{\Delta_0^q}{z}\left[\sum_k (1-f_k)\frac{\partial H_q}{\partial f_k}\phi_\nu(k,z) - (H_q(f)+1)\phi_\nu(q,z)\right] + \frac{1}{2}\sum_{k,k'}(1-f_k)(1-f_{k'})L_z[\phi_\nu(k,t)\phi_\nu(k',t)]\frac{\partial^2 H_q}{\partial f_k \partial f_{k'}} - \frac{C_q}{1-f_q} z\phi_\nu^2(z) = 0 \tag{43}$$

with

$$\Delta_0^q = H_q(f) - \frac{f_q}{1-f_q} . \tag{44}$$

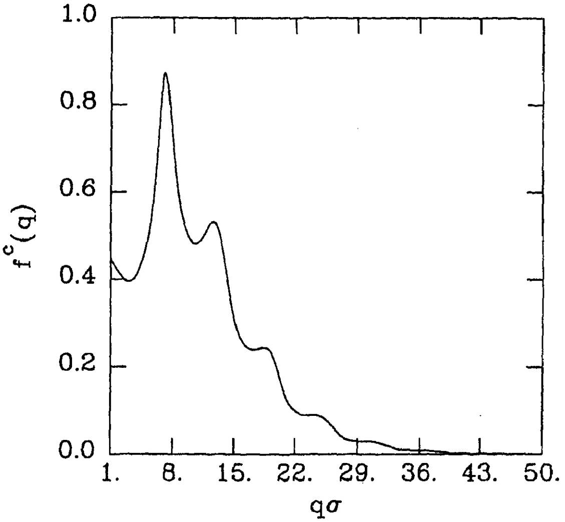


Figure 2. The long time limit of the density correlation function f_q, computed from the simple model at the density $n^* = 0.99$.

As in the other case here also we determine the f_q's by choosing the term linear in ϕ_ν to be vanishing. This gives rise to the condition that

$$H_q + 1 = \frac{C_q}{1 - f_q}, \tag{45}$$

where C_q are constants. Consequently, with this choice we can write

$$\Delta_0^q = \frac{C_q}{1 - f}. \tag{46}$$

For the ideal transition point one has the condition $\Delta_0^q = 0$ i.e. $C_q = 1$. Thus we obtain the following condition for f_q's,

$$f_q = \frac{H_q}{H_q + 1}. \tag{47}$$

In figure 2 we show the solution for f_q that is obtained using the model described by the kernel in equation (15). For the density $n^* \equiv n\sigma^3$ below 0.99 f_qs' all are zero showing that the dynamic instability signifying transition to the nonergodic phase takes place at this density. In obtaining this equation we get the static structure factor or the direct correlation function that appear in the mode coupling integrals using the Percus Yevick (PY) solution [10] for the hard sphere system. For the points away from the transition we see that with $\Delta_0^q > 0$ the function $\phi_\nu(q, z) \sim (\Delta_0^q)^{1/2}/z$. Thus the nonergodicity parameter giving the strength of the $1/z$ pole in the density correlation function has the behavior

$$f_q = f_q^0 + c_q \epsilon^{1/2} + O(\epsilon), \tag{48}$$

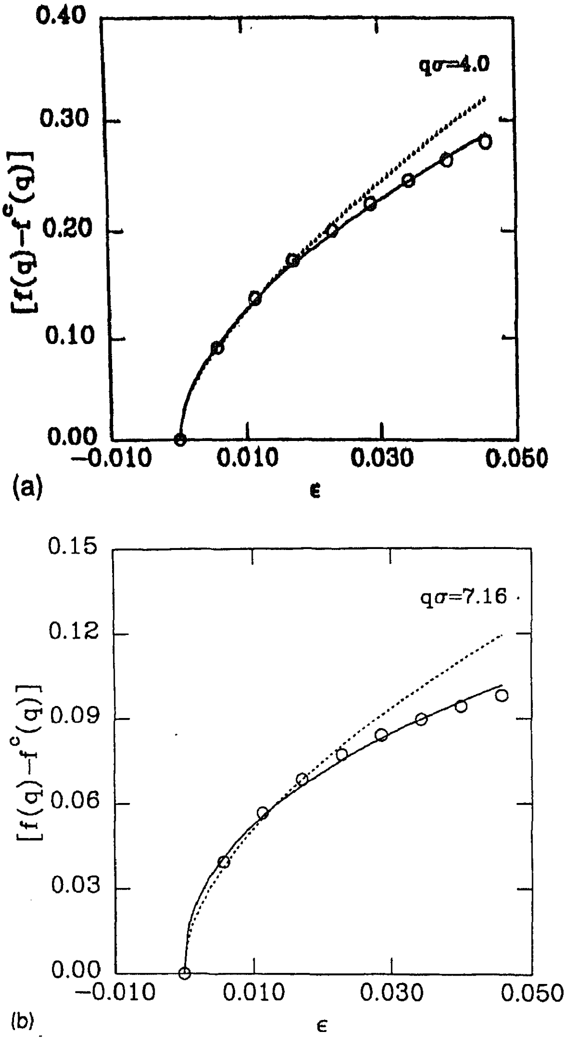


Figure 3. The nonergodicity parameter $[f_q - f_q^0]$ vs. ϵ for **(a)** $q\sigma = 4.00$; **(b)** $q\sigma = 7.16$. The dashed line is for the exponent $\frac{1}{2}$ and the solid line for a q dependent exponent.

where ϵ denotes the distance from the ideal transition point. Such a behavior was also found by numerical solution of the integral equations (47) giving f_q's. In figures 3(a) and 3(b) we show the fit for a power law with exponent 1/2 (dashed line) and a wave vector dependent exponent (solid line) for two different values of the wave vector $q\sigma = 4.00$ and 7.16 respectively. For distances (in temperature or density) from the transition which are very small the square root law seems to work but for large ϵ this seems to break down. This square root law with the non ergodicity parameter is indeed seen in a polymeric liquid [17] where the transition point is found to be T_0 =216 K, about 35 K above the phenomenological glass transition point. This observation has been viewed as important solution of the mode coupling equations.

4. The self-consistent model with the ergodicity restoring mechanism

If the quantity $\gamma(q,z)$ is maintained then the system shows ergodicity over long time scales. In ref. [5] it was demonstrated that the quantity $\gamma(q,z)$ can be obtained in terms of the hydrodynamic correlation functions by analyzing the self energy matrix element $\Sigma_{\hat{v}_i\hat{v}_j}$ introduced in (10) in the form

$$\gamma(q,t) = \int \frac{\mathrm{d}\mathbf{k}}{\rho_0^2(2\pi)^3}\left([u_1^2 G_{gg}^{\mathrm{L}}(\mathbf{q}-\mathbf{k},t) + (1-u_1^2)G_{gg}^{\mathrm{T}}(\mathbf{q}-\mathbf{k},t)]G_{\rho\rho}(\mathbf{k},t)\right.$$
$$\left. + uu_1 G_{g\rho}(\mathbf{k},t)G_{g\rho}(\mathbf{q}-\mathbf{k},t)\right) \tag{49}$$

with definition $u \equiv \hat{k}\cdot\hat{q}$ and $u_1 = \hat{q}(\mathbf{q}-\mathbf{k})/|\mathbf{q}-\mathbf{k}|$ and the superscripts L and T respectively refer to longitudinal and transverse parts of the corresponding quantities in the isotropic fluid. This self consistent expression for γ was also used [12] for obtaining a close set of functional equations for the density and current correlation functions. The time scales of relaxation that followed from numerical solution of these equations demonstrated good agreement with computer simulation results [19]. Indeed if γ would self-consistently reduce to very small values, the supercooled dynamics is pushed to very long times. Hence the exact form of the kernel $\gamma(q,z)$ representing the cutoff mechanism for eliminating the glass transition singularity is important. In ref. [12] this issue was considered through a self-consistent solution of (39) and (49). In figure 4 we show the behavior of the static longitudinal viscosity $D_{\mathrm{L}}(0,0) = \bar{\Gamma}$ with the reduced density in a hard sphere system. Initially the viscosity shows a power law increase (with exponent close to 2) but for higher densities the sharp transition is cutoff. This work demonstrated that the relaxation time increases by two to three orders of magnitude showing a change in the dynamics but it did not give rise to any diverging time scales around this mode coupling singularity.

The question regarding the behavior of the quantity γ i.e. how small it can be pushed to produce extremely slow relaxation has also been investigated. This is done by using a lower cutoff time in the computation of the mode coupling effects. The argument for this cutoff time is as follows: the mode coupling contribution to the memory function computes effects due to correlated collisions or cooperative dynamics in the supercooled liquid. The long time and short time contributions to the viscosity can be separated according to

$$D_{\mathrm{L}}^{mc}(q,z) = \int_0^{t_0} \mathrm{d}t\, \mathrm{e}^{izt} D_{\mathrm{L}}^{mc}(q,t) + \int_{t_0}^{\infty} \mathrm{d}t\, \mathrm{e}^{izt} D_{\mathrm{L}}^{mc}(q,t), \tag{50}$$

where the time t_0 sets the scale beyond which cooperative dynamics is effective. The first integral on the right side of (50) represents short time dynamics and is taken into account

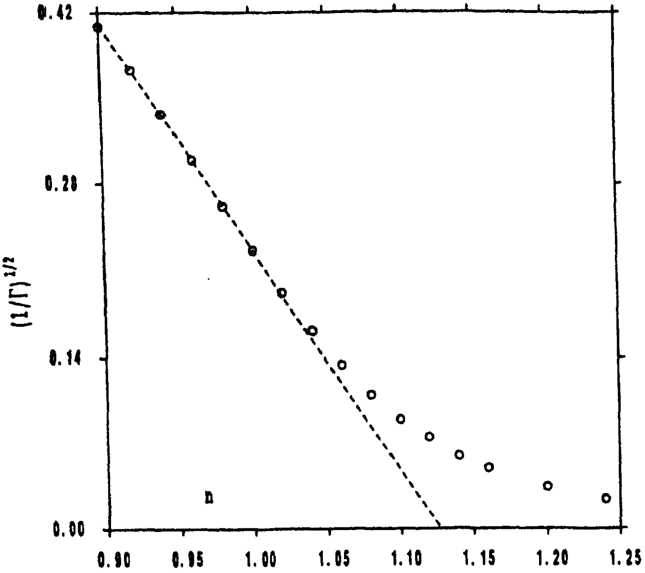


Figure 4. Square root of the inverse of the normalized longitudinal viscosity $\bar{\Gamma}$ vs. density. The dashed line shows the power law fit for intermediate densities.

through the bare transport coefficients. The mode coupling contribution to the viscosity is to a good approximation given by approximating it by the second integral in (50) and ignoring the first one since the bare transport coefficient already takes short time dynamics into account. For supercooled densities the same cutoff time is included in the expression for the kernel γ [20, 13]. The cutoff time t_0 in the theory can be estimated by fitting the value of the integrand with the computer simulation results on transport coefficients [15, 21]. For a hard sphere system good agreement with simulation data near freezing densities were obtained with $t_0 = 10t_{\rm E}$ where $t_{\rm E}$ is the Enskog time [22] referring to the time scale for short time dynamics for the hard sphere liquid. For supercooled densities however, the mode coupling contribution to the viscosity is not very sensitive to t_0, but γ does get influenced by the cutoff time. In a number of recent works [23] the experimental data for relaxation in supercooled systems has been used to obtain an estimation of the cutoff function. In this respect the computation of the cutoff function from a fully self-consistent approach is relevant. It has been observed that even if the short time part of the dynamics is excluded from the mode coupling contribution through t_0, the self-consistent calculation does not produce any drastic change in the relaxation time as is seen in the strong liquids. In a number of recent works [23] the cutoff mechanism, smoothing off the ideal glass transition in the framework of mode coupling theory, has been incorporated with the use of experimental data. The strength of the cutoff function is estimated there through an analysis of the data. In ref. [16] Kim investigated the behavior of the cutoff function in a schematic model where all q dependences are dropped from the model and obtained an expression of the form

$$\gamma = \int_{t_0}^{\infty} \mathrm{d}t [\dot{\phi}(t)]^2. \tag{51}$$

It was found [16] that if t_0 was chosen sufficiently large ($t_0 \equiv 5\Omega_0{}^{-1}$), γ is greatly suppressed. Therefore the suppression of γ is obtained once the cutoff time exceeds the microscopic value. In the fully q dependent theory we see a qualitatively similar but less enhanced effect. With reasonable approximations for t_0 relevant for the short time dynamics, does not give rise to the very strong enhancement of the viscosity. It does indicate a slowing down of the relaxation over two to three orders of magnitudes over a temperature or density range higher than the glass transition temperature. This is similar to what has been termed in literature [24] as fragile liquids. It does not however give rise to, within sensible approximations, in the framework of a self consistent calculation the very sharp increase in viscosity seen in the strong liquids. Thus the basic results for the relaxation time does not change drastically with the introduction of the cutoff time t_0 in the framework of the fully self-consistent calculation.

5. Discussion

In this paper we discuss the nonlinear fluctuating hydrodynamic equations for compressible liquid, and how they were used for obtaining the mode coupling models for the supercooled liquid dynamics. The theory demonstrated existence of a characteristics temperature T_0 higher than the calorimetric glass transition temperature T_G such that within a narrow temperature range around T_0 there is a freezing out of the large scale structural rearrangements involving collective motion of many molecules. At this temperature there is a qualitative change in the collective dynamics in the liquid although there is no sharp glass transition characterized by a diverging viscosity. The temperature T_0 [27, 7, 23] is a signature of the dynamic singularity due to mode coupling effects. In a careful analysis [5] of the NFH equations it was demonstrated that ergodicity is maintained at all densities. In the Das and Mazenko analysis the role of the $1/\rho$ nonlinearities appearing in the equations were implemented in the field theoretic analysis through the constraint $\mathbf{g} = \rho\mathbf{v}$. The cutoff mechanism responsible for the absence of the sharp transition is a direct consequence of this [29]. Ergodicity was also demonstrated [25] in the asymptotic dynamics obtained in similar mode coupling models obtained from microscopic approaches. In a subsequent work Schimitz De and Dufty [26] (hereafter referred to as SDD) has also considered the equations of nonlinear fluctuating hydrodynamics to obtain a self consistent mode coupling theory for supercooled liquids extending it to short wavelengths. The analysis presented by these authors demonstrate the absence of a sharp transition to an ideal glassy phase similar to the earlier work by Das and Mazenko. In the work by SDD the role of the nonlinearities in the fluctuating hydrodynamic equations are investigated with the underlying microscopic dynamics being constrained by the detailed balance condition. The authors show that special nonlinearities of density ρ and momentum field $\mathbf{g}$ which appear in the continuity equation to maintain detailed balance, do eliminate a complete structural arrest that would have occurred if only coupling of density fluctuations were considered. In a recent work [28] it was demonstrated that both the works gives the identical result for the final relaxation process. The absence of the sharp glass transition in the mode coupling models is linked to the fact that with the increase of density the kernel γ [5] does not self consistently reduce to zero and the density correlation function decays to zero in the long time limit.

Indeed such an effect is also observed in a wide range of systems showing a qualitative change in dynamics of the supercooled liquid around a characteristics temperature higher than the calorimetric glass transition temperature although there is no sharp transition to an ideal glassy phase. γ however gets small, indicating a two/three orders of magnitude rise in the value of transport coefficients. In ref. [30] the authors have used a sophisticated form of scaling in α-relaxation of supercooled liquids. From the measurements of the dielectric susceptibility of a variety of glass forming liquids they observed that the data for all the sample liquids studied can be scaled so that they fall on top of one another over 13 decades of frequency. The curve is described by two parameters, the peak frequency ν_p and the normalized width W (with respect to the Debye width $= 1.14$) of the imaginary part of the dielectric susceptibility. While the observed scaling does fit very well to the stretched exponential form for low frequencies it deviates significantly from the stretched exponential form on the high frequency side. In a subsequent analysis by Kim and Mazenko [13] demonstrated the existence of a cross over temperature in the data of Nagel *et al* by plotting inverse of the peak frequency ν_p with temperature. The high frequency data again fits well to the form $(T - T_0)^2$ and $T_0 = 270\,\mathrm{K}$ for salol. These authors have also argued that the high frequency tail observed in Nagel's scaling curve is a realization of the von-Schweidler relaxation in the mode coupling theory if one assumes a relation between the exponent β of stretched exponential relaxation with the exponent b of von-Schweidler relaxation. Such a relation does not however follow from the mode coupling models obtained from simple nonlinear hydrodynamic models. Thus while the mode coupling models has led to considerable progress in understanding the sequence of time scales associated with relaxation near glass transition, there remain still significant questions for further research.

Acknowledgements

Saha Institute of Nuclear Physics is gratefully acknowledged for hospitality on a visit during the preparation of this manuscript in the summer 1996. Amit Dutta is acknowledged for his kind help in preparing the manuscript.

References

[1] P Das, G F Mazenko, S Ramaswamy and J Toner, *Phys. Rev. Lett.* **54**, 118 (1985)
[2] T R Kirkpatrick, *Phys. Rev.* **A31**, 939 (1985)
[3] E Leutheusser, *Phys. Rev.* **A29**, 2765 (1984)
[4] U Bengtzelius, W Götze and A Sjölander, *J. Phys.* **C17**, 5915 (1984)
[5] S P Das and G F Mazenko, *Phys. Rev.* **A34**, 2265 (1986)
[6] P Taborek, R N Kleiman and D J Bishop, *Phys. Rev.* **B34**, 1835 (1986)
[7] W Götze, in *Liquids, Freezing and the Glass Transition* edited by D Levisque, J P Hansen and J Zinn-Justin (Elsevier, New York, 1991)
[8] R Zwanzig and M Bixon, *Phys. Rev.* **A2**, 2005 (1970)
H Mori, *Prog. Theor. Phys.* **49**, 1516 (1973)
[9] J Langer and L Turski, *Phys. Rev.* **A8**, 3230 (1973)
[10] J P Hansen and J R Mcdonald, *Theory of Simple Liquids* (Academic, London, 1976)
[11] P C Martin, E D Siggia and H A Rose, *Phys. Rev.* **A8**, 423 (1973)
U Deker and F Haake, *Phys. Rev.* **A11**, 2043 (1975)

U Deker, *Phys. Rev.* **A19**, 846 (1979)
For the functional integral formulation of the theory used here, see R Bausch, H Janssen and H Wagner, *Z. Phys.* **B24**, 113 (1976)
H J Janssen, in *Dynamical Critical Phenomena and Related Topics* edited by C P Enz (Springer-Verlag, New York, 1979)
C De Dominicis and L Pelti, *Phys. Rev.* **B18**, 353 (1978)
R V Jensen, *J. Stat. Phys.* **25**, 183 (1981)
[12] S P Das, *Phys. Rev.* **A42**, 6116 (1990)
[13] B Kim and G F Mazenko, *Adv. Chem. Phys.* **78**, 129 (1990)
[14] W Götze, *Z. Phys.* **B56**, 139 (1984); **60**, 195 (1985)
[15] S P Das and J W Dufty, *Phys. Rev.* **A46**, 6371 (1992)
[16] B Kim, *Phys. Rev.* **A46**, 1992 (1992)
[17] D Richter, B Frick and B Farago, *Phys. Rev. Lett.* **61**, 2465 (1988)
B Frick, B Farago and D Richter, *Phys. Rev. Lett.* **64**, 2921 (1990)
[18] S P Das, *J. Chem. Phys.* **98**, 3328 (1993)
[19] J J Ullo and S Yip, *Phys. Rev. Lett.* **54**, 1509 (1985); *Phys. Rev.* **A39**, 5877 (1989)
[20] L Sjögren, *Z. Phys.* **B79**, 5 (1990)
[21] I M de Schepper, A F E M Haffmans and H van Beijeren, *Phys. Rev. Lett.* **57**, 1715 (1986)
[22] J P Boon and S Yip, *Molecular Hydrodynamics* (McGraw–Hill, New York, 1979)
[23] M Fuchs, W Götze, S Hildebrand and A Latz, *J. Phys. Condens. Matter* **4**, 7709 (1992)
H Z Cummins, W Du, M Fuchs, W Götze, S Hildebrand, A Latz, G Li and N Tao, *Phys. Rev.* **E47**, 4223 (1993)
[24] C A Angell, *Proc. Workshop on Relaxation Processes*, Blacksburg, Va., July, 1983
[25] W Götze and L Sjoren, *Z. Phys.* **B65**, 415 (1987)
[26] R Schmitz, J W Dufty and P De, *Phys. Rev. Lett.* **71**, 2066 (1993)
[27] J P Hansen, *Physica* **A201**, 138 (1993)
[28] S P Das, *Phys. Rev.* **E54**, 463 (1996)
[29] G F Mazenko and J Yeo, *J. Stat. Phys.* **74**, 1017 (1994)
[30] P K Dixon, L Wu, S R Nagel, B D Williams and J P Carini, *Phys. Rev. Lett.* **65**, 1108 (1990)

G Ambika Department of Physics, Maharajas College, Cochin 682 011, India

R E Amritkar Department of Physics, University of Pune, Pune 411 007, India
Email: rea@chaos.physics.unipune.ernet.in

G Ananthakrishna Materials Research Centre, Indian Institute of Science, Bangalore 560 012, India
Email: garani@mrc.iisc.ernet.in

Balakrishnan Radha The Institute of Mathematical Sciences, Madras 600 113, India
Email: radha@imsc.ernet.in

V Balakrishnan Department of Physics, Indian Institute of Technology, Madras 600 036, India
Email: phy5@iitm.ernet.in

Bala Sundaram Department of Mathematics, CSI-CUNY, 2800, Victory Boulevard, Staten Island, NY 10314, USA
Email: bas@george.ph.utexas.edu

Bambah A Bindu School of Physics, University of Hyderabad, Hyderabad 500 046, India

J R Bhatt Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India

J K Bhattacharjee Department of Theoretical Physics, Indian Association for the Cultivation of Science, Calcutta 700 032, India
Email: tpjkb@iacs.ernet.in

A Bhowal Saha Institute of Nuclear Physics, Calcutta 700 064, India
Email: avijit@hp2.saha.ernet.in

Biswas Debabrata Theoretical Physics Division, Bhabha Atomic Research Centre, Bombay 400 085, India
Email: krsrini@magnum.barct1.ernet.in

A B Chhabra Chemical Bank, New York, NY 10021, USA

Das K Parichay Indian Institute of Chemical Technology, Uppal Road, Hyderabad 500 007, India

Das P Shankar School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067, India
Email: shankar@jnuniv.ernet.in

Dhar K Sujan Department of Physics, Indian Institute of Science, Bangalore 560 012, India

C Fressengeas Laboratoire de Physique et Mécanique des Matériaux, URA 1215 CNRS, Institut Supérieur de Génie Mécanique et Productique, Université de Metz, Ile du Saulcy, 57045 - Metz Cedex 01, France

Gadre D Sangeeta Department of Physics and Astrophysics, University of Delhi, Delhi 110 007, India

K Ganesan Department of Applied Mathematics and Theoretical Physics, The Queen's University of Belfast, Belfast BT7 1NN, United Kingdom
Email: ganesan@jaguar.am.qub.ac.uk

Gangal D Anil Department of Physics, University of Pune, Pune 411 007, India
Email: gangal@chaos.physics.unipune.ernet.in

Gaspard Pierre Faculté des Sciences and Center for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Campus Plaine C.P. 231, Boulevard du Triomphe, B-1050 Brussels, Belgium

R Gebarowski Department of Applied Mathematics and Theoretical Physics, The Queen's University of Belfast, Belfast BT7 1NN, United Kingdom

B N Goswami Centre for Atmospheric and Oceanic Sciences, Indian Institute of Science, Bangalore 560 012, India
Email: goswamy@caos.iisc.ernet.in

Gupte Neelima Department of Physics, Indian Institute of Technology, Madras 600 036, India
Email: gupte@imsc.ernet.in

Jain R Sudhir Theoretical Physics Division, Bhabha Atomic Research Centre, Bombay 400 085, India

P K Kaw Institute of Plasma Research, Bhat, Gandhinagar 382 424, India
Email: kaw@plasma.ernet.in

Ketoja A Jukka Department of Mathematics, P.O. Box 4, FIN-00014, University of Helsinki, Finland

Khare Avinash Institute of Physics, Sachivalaya Marg, Bhubaneswar 751 005, India
Email: khare@iopb.ernet.in

Kolwankar M Kiran Department of Physics, University of Pune, Pune 411 007, India
Email: kirkol@physics.unipune.ernet.in

A Kudrolli Physics Department, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA
Email: akudroll@haverford.edu

B D Kulkarni Chemical Engineering Division, National Chemical Laboratory, Pune 411 008, India
Email: bdk@ncl.ernet.in

A Lahiri Vidyasagar Evening College, Calcutta 700 006, India

M Lakshmanan Centre for Nonlinear Dynamics, Department of Physics, Bharathidasan University, Tiruchirapalli 620 024, India
Email: lakshman@cnld.rect.ernet.in

S Lakshmibala Department of Physics, Indian Institute of Technology, Madras 600 036, India

Lakshminarayan Arul Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India
Email: arul@prl.ernet.in

J M Luck C.E.A. Saclay, Service de Physique Théorique, 91191 Gif-sur-Yvette cedex, France

Mahapatra Susanta Department of Chemistry, Indian Institute of Technology, Kanpur 208 016, India

Mehra Vishal School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067, India

Mehta Anita S.N. Bose National Centre for Basic Sciences, JD Block, Sector III, Salt Lake City, Calcutta 700 091, India
Email: anita@bose.ernet.in

Mehta Madan Lal Centre d'Études Nucleaires de Saclay, 91191 Gif-sur-Yvette Cedex, France
Email: mehta@amoco.saclay.fr

C Mukku School of Mathematics and Computer/Information Sciences, University of Hyderabad, Hyderabad 500 046 India

V M Nandakumaran International School of Photonics, Cochin University of Science and Technology, Cochin 682 022, India

Nayak K Saroj Department of Physics, Virginia Commonwealth University, Richmond, VA 23284, USA
Email: snayak@oscar.vcu.edu

C Nicolis Institut Royal Météorologique de Belgique, 1180 Bruxelles, Belgium

G Nicolis Centre for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, C.P. 231, 1050 Bruxelles, Belgium

S J Noronha Materials Research Centre, Indian Institute of Science, Bangalore 560 012, India

D J Olinger Mechanical Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA

Pal Santanu Variable Energy Cyclotron Centre, 1/AF Bidhan Nagar, Calcutta 700 064, India
E-mail: santanu@veccal.ernet.in
Pande Ashwin Department of Physics, Indian Institute of Science, Bangalore 560 012, India
Pandit Rahul Department of Physics, Indian Institute of Science, Bangalore 560 012, India
Email: rahul@physics.iisc.ernet.in
Parekh Nita Chemical Engineering Division, National Chemical Laboratory, Pune 411 008,
India
J C Parikh Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India
K Porsezian Institut für Theoretische Physik, Universität Hannover, 30167 Hannover,
Germany
Email: ponz@itp.uni-hannover.de
Centre for Laser Technology, Department of Physics, Anna University, Madras 600 025, India
Puri Sanjay School of Physical Sciences, Jawaharlal Nehru University, New Delhi 110 067,
India
Email: puri@jnuniv.ernet.in
L Quaouire Laboratoire de Physique et Mécanique des Matériaux, URA 1215 CNRS,
Institut Supérieur de Génie Mécanique et Productique, Université de Metz, Ile du Saulcy,
57045 - Metz Cedex 01, France
R Radha Centre for Nonlinear Dynamics, Department of Physics, Bharathidasan University,
Tiruchirapalli 620 024, India
S Rajasekar Department of Physics, Manonmmaniam Sundaranar University, Tirunelveli
627 002, India
Ramaswamy Ramakrishna School of Physical Sciences, Jawaharlal Nehru University,
New Delhi 110 067, India
Email: rama@jnuniv.ernet.in
Rangarajan Govindan Department of Mathematics and Centre for Theoretical Studies,
Indian Institute of Science, Bangalore 560 012, India
Email: rangaraj@math.iisc.ernet.in
V Ravi Kumar Chemical Engineering Division, National Chemical Laboratory, Pune 411 008,
India
Email: ravi@ncl.ernet.in
Roy Manojit Department of Physics, University of Pune, Pune 411 007, India
Email: mroy@physics.unipune.ernet.in
T K Roy Saha Institute of Nuclear Physics, Calcutta 700 064, India
Sain Anirban Department of Physics, Indian Institute of Science, Bangalore 560 012, India
M S Santhanam Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India
N Sathyamurthy Department of Chemistry, Indian Institute of Technology, Kanpur 208 016,
India
Email: nsath@iitk.ernet.in
Satija I Indubala Department of Physics and Institute of Computational Sciences and
Informatics, George Mason University, Fairfax, VA 22030, USA
Email: isatija@sitar.gmu.edu
S V M Satyanarayana Theroretical Studies Section, Materials Science Division, Indira
Gandhi Centre for Atomic Research, Kalpakkam 603 102, India
J Segar Department of Theoretical Physics, University of Madras, Guindy Campus,
Madras 600 025, India
A Sen Institute of Plasma Research, Bhat, Gandhinagar 382 428, India
Email: abhijit@plasma.ernet.in
S Sengupta Institute of Plasma Research, Bhat, Gandhinagar 382 424, India
Sharma Ashutosh Department of Physics, University of Pune, Pune 411 007, India
V B Sheorey Physical Research Laboratory, Navrangpura, Ahmedabad 380 009, India
Email: sheorey@prl.ernet.in

Sinha Somdatta Centre for Cellular & Molecular Biology, Uppal Road, Hyderabad 500 007, India
Email: sinha@ccmb.uunet.in

Sinha Sudeshna The Institute of Mathematical Sciences, Madras 600 113, India
Email: sudeshna@imsc.ernet.in

K R Sreenivasan Mason Laboratory, Yale University, New Haven, CT 06520, USA
Email: krs@kolmogorov.eng.yale.edu

S Sridhar Physics Department, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA
Email: sridhar@neu.edu

V Sridhar Theoretical Studies Section, Materials Science Division, Indira Gandhi Centre for Atomic Research, Kalpakkam 603 102, India

M S Sriram Department of Theoretical Physics, University of Madras, Guindy Campus, Madras 600 025, India

M C Valsakumar Theoretical Studies Section, Materials Science Division, Indira Gandhi Centre for Atomic Research, Kalpakkam 603 102, India

V S Varma Department of Physics and Astrophysics, University of Delhi, Delhi 110 007, India